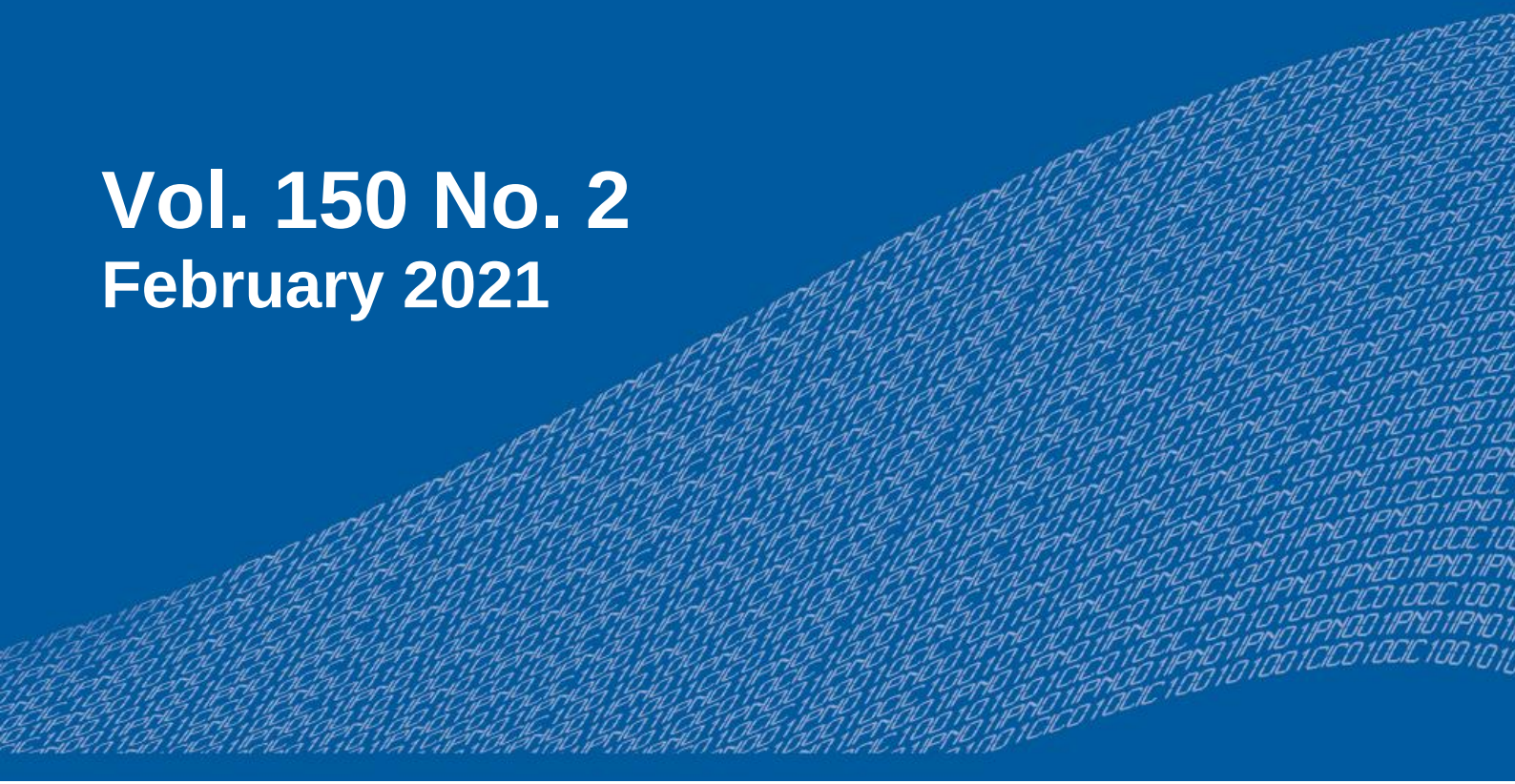
EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA

Instituto Politécnico Nacional
"La Técnica al Servicio de la Patria"

Research in Computing Science

ISSN: 1870-4069



Vol. 150 No. 2
February 2021

Research in Computing Science

RESEARCH IN COMPUTING SCIENCE, *Año 21, Volumen 150, No. 2, 05 de Febrero de 2021, es una publicación mensual, editada por el Instituto Politécnico Nacional, a través del Centro de Investigación en Computación. Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738, Ciudad de México, Tel. 57 29 60 00, ext. 56571. https:// www.rcs.cic.ipn.mx. Editor responsable: Dr. Grigori Sidorov. Reserva de Derechos al Uso Exclusivo del Título No. 04-2019-082310242100-203 ISSN: en trámite, ambos otorgados por el Instituto Nacional del Derecho de Autor. Responsable de la última actualización de este número: el Centro de Investigación en Computación, Dr. Grigori Sidorov, Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738. Fecha de última modificación 05 de Febrero de 2021.*

RESEARCH IN COMPUTING SCIENCE, *Year 21, Volume 150, No. 2, February 05, 2021, is a monthly publication edited by the National Polytechnic Institute through the Center for Computing Research. Av. Juan de Dios Bátiz S/N, Esq. Miguel Othón de Mendizábal, Nueva Industrial Vallejo, C.P. 07738, Mexico City, Tel. 57 29 60 00, ext. 56571. https://www.rcs.cic.ipn.mx. Editor in charge: Dr. Grigori Sidorov. Reservation of Exclusive Use Rights of Title No. 04-2019-082310242100-203. ISSN: pending, both granted by the National Copyright Institute. Responsible for the latest update of this issue: the Computer Research Center, Dr. Grigori Sidorov, Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738. Last modified on February 05, 2021.*

Volume 150(2)

Hybrid Intelligent Systems

Martín Montes Rivera
Carlos Alberto Ochoa Ortiz (eds.)

Instituto Politécnico Nacional,
Centro de Investigación en Computación, México 2021

ISSN: en trámite

Formerly ISSNs: 1870-4069, 1665-9899

Instituto Politécnico Nacional (IPN)
Centro de Investigación en Computación (CIC)
Av. Juan de Dios Bátiz s/n esq. M. Othón de Mendizábal
Unidad Profesional "Adolfo López Mateos", Zacatenco
07738, México D.F., México

http://www.rcs.cic.ipn.mx
http://www.ipn.mx
http://www.cic.ipn.mx

Indexed in LATINDEX, DBLP and Periodica

Electronic edition

ISSN 1870-4069

Table of Contents

Page

ISSN 1870-4069

Intelligent Hybrid Decision Making System Focused on Recommendations for the Treatment of Heterogeneous Seasonal Citrus Fields

Antonio Romero[1], Eddy Sánchez-DelaCruz[1], Alberto Ochoa[2]

[1] Instituto Tecnológico Superior de Misantla, Veracruz, Mexico

[2] Universidad Autónoma de Ciudad Juárez, Mexico

romeroa_cv123@hotmail.com, esanchezd@misantla.tecnm.mx, alberto.ochoa@uacj.mx

Abstract. The implementation of an intelligent drone that replaces one of the main tasks of farmers, direct visual inspection, in order to reduce future damage to citrus production generated by bird pests, where the use of a convolutional neural network will serve to extract characteristics and generate a classification model capable of distinguishing between healthy foliage of a tree and that which is not, Consequently, it presents some kind of pest produced by the birds taking into account the coloration of the leaves, as well as, to extract the coloration of the leaves that present diseases, with the purpose of optimizing and improving the activities immersed in the cultivation of the citrus fruits. In addition, to reduce the damage of pests of birds, which will significantly increase productivity in the agricultural sector annually, therefore, through the implementation of unmanned aerial vehicles, provide visual recognition to reduce the indirect effects generated by the type of pests and thus optimize the activities and techniques of citrus cultivation, carrying out the incorporation of the image repository for decision making created.

Keywords: Smart drone, citrus cultivation, bird pests, convolutional neural network.

1 Introduction

One of the main problems that farmers have to face is pests in their crops that generate great losses and threaten food security, as well as economic problems for the farmer due to losses. Currently, the technique used to detect bird attacks on crops is a traditional method of control in the fields, direct visual inspection, which consists of personally supervising the planting, that is, the farmer has to go to the area of interest in order to watch and monitor the crops, this method is slow and is not applicable to large areas of land (Bokolonga et al., 2016). However, there are other methods that experience efficiency, but at a higher cost, therefore, the detection of attacks on crops

is suitable for viewing from the altitude with the help of an application that allows easy use and detection of the damaged crop.

On the other hand, drone technology provides a solution to the conflicts or difficulties that currently exist for a farmer because they allow for a more extensive and complete expansion of inspections, thus helping to work in the agricultural sector, in addition, manual work is reduced with the implementation of this collaboration, since it benefits them in the field activities due to the fact that the unmanned aerial vehicle becomes the eyes of the citrus grower in order to make this task easier, since with the help of the cameras it captures images showing if the crop has pests and showing the state of the plantation and improving the crop control techniques (Vatalaro et al., 2016).

On the other hand, there are a variety of artificial intelligence techniques that can be implemented and that can help improve the agricultural sector, so we worked with a Neuronal Convolutional Network, which was trained with images of citrus crops to achieve greater accuracy in pest detection and that after this, this classification model was intended for a mobile application for greater control and also to allow better management (De Rango et al., 2019).

In contrast, once the detection of the disease caused by bird pests has been made, this analysis serves as a repository for a new classification, which is stored directly in the cloud. The different diseases that this pest causes to citrus plantations were analyzed and recognized by expert growers in the area of Misantla, obtaining a coincidence of recognition that exceeds 98% of identification in which the coloring of the foliage is taken into account, as well as the citrus fruit.

1.1 Object Classifier and Object Detector

Firstly, it is important to know how to differentiate between an object classifier and an object detector. An image classifier is an algorithm that is dedicated to classifying images within a specific category. For example, a group of images of a crop with pests were assigned and this only mentions that the predominant object in the photo is the particular damage, while an object detector is an algorithm that is responsible for identifying various elements within the image and thus classifying them, for example, they receive an image containing a citrus crop and identify that there is a pest, bird, tree, etc. within the image (Gonzalez, 2007).

1.2 Artificial Neural Network

Artificial Neural Networks are part of Artificial Intelligence, they are also networks trained through the inputs obtained from external or internal scenarios in the system and these inputs are multiplied by randomly assigned weights, a neural network is an integration of various learning systems, that is why they have the ability to learn through previous training, the ANNs are programming objects that mimic the functioning of biological neurons (Rivas-Asanza et al., 2018).

On the other hand, NNAs have many advantages such as adaptive learning in which you learn to perform tasks from a set of data being that in the process of learning these data are represented as inputs and weights, In addition, self-organization as you can create your own organization or representation of information received, the neural networks self-organize their information they receive during the learning of the

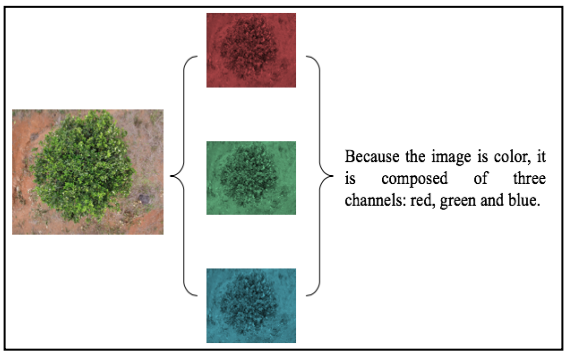


Fig. 1. RGB example.

operation using the mathematical methods Adeline, Madeline and Perceptron among others. Additionally, tolerance to partial failures knowing that the partial destruction of the network damages its operation but does not destroy it completely. This is due to the redundancy of the information contained, that is, this leads to the information not being lost since it works like the human body. Similarly, their operation is in real time, as they can be carried out by computers or special hardware devices to take advantage of the capacity of the NNAs (Matich, 2001).

Convolutional Neural Network

CNNs are similar to ordinary neural networks such as the multi-layer perceptron, they are composed of neurons that have weights and biases that they can learn. Each neuron receives some input, performs a scalar product and then applies a trigger function. What distinguishes CNNs is that they explicitly assume that the inputs are images, which allows us to encode certain properties in the architecture; favoring efficiency gains and reducing the amount of parameters in the network. CNNs even solved the problem that ordinary neural networks do not scale well for high definition images. CNNs work by consecutively modeling small pieces of information, and then mixing the information in the deeper layers of the network, so they are able to model complex variations and behaviors giving fairly accurate predictions (Briega, 2016).

The CNN is a type of Artificial Neural Network with supervised learning that processes its layers imitating the visual cortex of the human eye to identify different characteristics in the inputs that ultimately make it able to identify objects and "see". For this, CNN contains several specialized hidden layers with a hierarchy: this means that the first layers can detect lines, curves and specialize until they reach deeper layers that recognize complex shapes such as a face or the silhouette of an animal (Na8, 2018).

Pixels and Neurons

To start, the network takes the pixels in the image as input. Here you have an image with 4000×3000 pixels high and wide, that's equivalent to 12000000 neurons. And that's in case you only have 1 color (grayscale). If you had a color image, you would need 3 RGB channels (network, green, blue) and then you would use 4000x3000x3 = 36000000 neurons as input.

Fig. 2. Persian Lemon tree.

Before feeding the grid, it is convenient to normalize the values as input. The colors of the pixels have values ranging from 0 to 255, a transformation of each pixel will be carried out: "value/255" and there will always be a value between 0 and 1. This range, due to the operation of convolutional networks, makes it possible to identify the characteristics for each vector incorrectly. To correct this, it is convenient to previously normalize the image, in a process known as zero mean and unit variance normalization. To summarize, it refers to a 3-dimensional matrix where each dimension represents a layer of color, and each of these layers has values ranging from 0 to 255, and represent the intensity of that color, where the computer to display takes a value and draws a pixel of the resulting color of the mixture.

2 Choice of the Crop to be Analized

Below, you will find statistical information on agricultural production in Misantla, Ver. on an annual basis in the products it generates, the information offered is on area sown, area harvested and the value of production, for seasonal crops in the area (SEDARPA, 2017).

The study includes crops with better production in the municipality, which are ordered in a descending manner, that is, from more to less production, which will help define the crop to be covered by the project.

It is possible to identify in the previous graph, four different types of products, which are: the lemon, the orange and the tangerine, being these first two elements those that produce more the region of Misantla, which gives to understand that as much the orange and the lemon are two of the fruits that but are cultivated, for that reason, they are taken like reference to be compared with the other products, in where the planted surface takes part, the production in weights, which is taken to determine the type of culture that will be to which the intelligent drone is implemented.

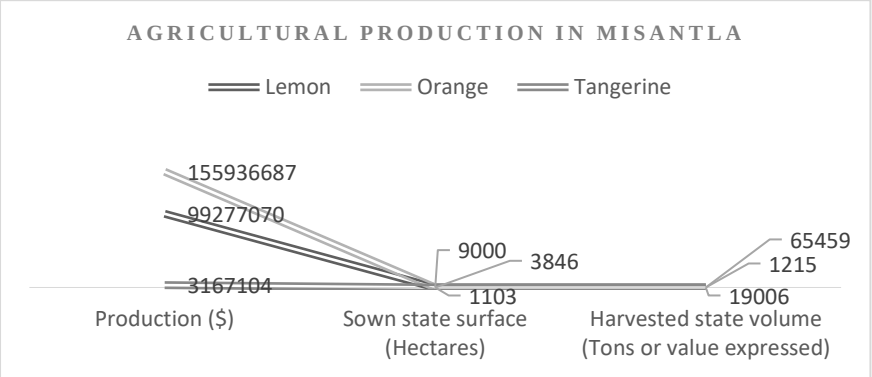


Fig. 3. Agriculture production.

Based on the data provided by the Secretariat of Agricultural, Rural and Fishing Development (SEDARPA) and taking as reference the analysis of the graph, it is concluded that the crop to work on the project and which will be implemented the use of drone, will be the citrus, since it presents a greater production in the city of Misantla, as well as a larger area of cultivation than the other fruits. On the other hand, it is proposed that in this type of citrus fruit is also affected by bird pests.

2.1 Process to be Automated

According to citrus growers in the Misantla area, Ver., the greening of citrus fruits can cause the shoots and foliage to die, giving them a scarce and untidy appearance, as well as not growing as much as they should or blooming at the wrong time of year, for example in summer or autumn instead of spring. Citrus greening, also known as Huanglongbing, is a bacterial infection transmitted by an insect called the Asian citrus psyllid (Mead, 2020).

In summary, the damage caused by HLB is as follows:

- Economic death of the plant.
- Severe fruit and leaf drop.
- Decrease in fruit weight.
- Decrease in the sugar level.
- Increase of the acidity level.
- Decrease in size and percentage of juice
- A young plant does not produce fruits.

The symptoms can be found throughout the year, however, are easier to see from September to March because in this season the color of the leaves changes and the pest becomes visible, the disease affects all parts of the tree's crown: leaves, twigs and fruits, as the disease progresses, it will cause the whole tree to decrease. In addition, leaf symptoms include mottled spots, yellow veins, clogged veins or green islands, yellow veins, clogged veins or green islands are not diagnostic alone, on the other hand mottled spotting is the best diagnostic symptom of green leaf. And spotting is a random pattern

Fig. 4. Leaf symptoms.

Fig. 5. Greening and nutrient deficiency problems.

of yellowing that occurs on leaves that is not the same on the right and left sides of the leaf (Florida, 2018).

A simple procedure to determine if the symptoms are the same in both halves of a leaf is the following, draw two circles in opposite halves of the leaf in order to appreciate if the pattern is the same in both circles as in the following figure. The figure above shows two types of problems, on the left side a green problem and on the right-side problems with nutrient deficiency with a line as suggested above.

In addition, the pest problem does not only affect the appearance of the leaf. The external appearance of the fruit can also be affected, whether it is deformed or looks small and green. Taste problems, a variation of salty and bitter. And the internal appearance may have aborted seeds, a yellow spot under the calyx button and/or a curved central core.

Finally, the identification of the problems that include all the tree, beginning to present/display yellow buds, death in its branches, delays in the growth, present/display flowering out of season and a remarkable general reduction of the tree.

2.2 Materials and Methods

Table 2 describes the characteristics of the laptop, as well as the drone and the system requirements of the two devices mentioned.

The following process was determined from the literature review that was done taking into account work related to digital image processing for the detection and analysis of images in precision agriculture, where the stress that plants present is taken into account to analyze the images.

First, the dataset was built up from images obtained by the DJI Mavic Mini drone. Once the dataset was built, a pre-processing of images was applied to filter and delimit regions of interest, followed by the application of segmentation to the images to identify the regions affected by the pests, in order to extract the characteristics and, here, a label was made to obtain the most relevant characteristics and identify them. Finally, to

Fig. 6. Fruit problems.

Fig. 7. Foliage problems.

achieve the classification, the crops with pest in order to later determine the damage they present.

2.3 Acquisition of the images

Were carried out 8 flights with a DJI mavic mini, the images taken by the drone are stored in jpeg format with a size of 5,438,160 bytes and dimensions of 4000 × 3000 pixels each, have a focal length of 4.49, are sectioned by folders according to the type of citrus, including 2 main, the Persian lemon and late orange valencia and a sub-folder within each type of citrus which refer to whether the crop has or not pest.

2.4 Development

The development of the application is the process of classifying images or recognizing them. In this process, an image is shown to the device's camera sensor and it will tell you which class the image belongs to. By means of image classification and training a deep learning model of convolutional neural network and the help of a google online tool, the model is exported to the TensorFlow lite version that is compatible with the Android device in order to implement it in the application.

To perform the training of the model the platform to perform the image classifier is teachable machine from google. At the end of the training, you can export the model of your choice. The model was exported to the Lite version of TensorFlow. Google's

Table 1. Material specifications.

Personal computer	
Model	MacBook Pro (13-inch, Early 2011)
System	macOS High Sierra
Type of System	64 bits
Procesador	2.3 GHz Intel Core i5
Graphics card	Intel HD Graphics 3000 512 MB
RAM	8 GB 1600 MHz DDR3
Hard disk drive	240GB SSD
Drone	
Model	DJI Mavic Mini
Take-off weight	Folded: 140 × 82 × 57 mm (length × width × height) Unfolded: 160 × 202 × 55 mm (length × width × height) Unfolded (with propellers): 245 × 290 × 55 mm (length × width × height)
Camera	
Sensor	1 / 2.3" CMOS Effective Pixels: 12PM/2.3" CMOS
ISO range	Video: 100 - 3200 (automatic) 100 - 3200 (manual) Photo: 100 - 1600 (automatic) 100 - 3200 (manual)
Photo size	4:3: 4000×3000 16:9: 4000×2250
Video resolution	FHD: 1920×1080 25/30/50/60 p 2.7K: 2720×1530 25/30 p
Maximum bit rate	40 Mbps
Photo formats	JPEG

Teachable Machine was used, which is an Artificial Intelligence platform that allows the user to train independently to a neural network, while the system can react to known images with sounds or GIF animations. At least 30 frames are required for the Teachable Machine to learn movement recognition. To start the training, the user must hold down the "train" button for a few seconds and make a certain gesture or show something to the computer (Ulasovich, 2016).

The tool allows you to immerse yourself in the definition of the model, algorithm and data processing as well as, focus only on the deployment of the model that is generated, the tool works in the browser, with a webcam or with files that are hosted locally or in Google Drive, and in a few minutes you can quickly understand how a model "learns" through a simple classification demonstration.

To start training the classification model, first different categories or classes were created to carry out the training and have it learn. Four classes were created as shown in figure 3, which are two types of citrus fruits with their respective healthy and pest variables. With the classes or labels defined, from the local storage the samples of each class are uploaded to the platform so that the training can begin. More than 200 images were uploaded with 4000 × 3000 resolution per class.

After having the classes ready, it is already possible to train the classification model. The model takes only a few minutes to train. Not only is the speed of training excellent, but you are given valuable dynamic metrics such as accuracy, training loss and dataset test with the observed class accuracy.

Fig. 8. Persian lemon harvest.

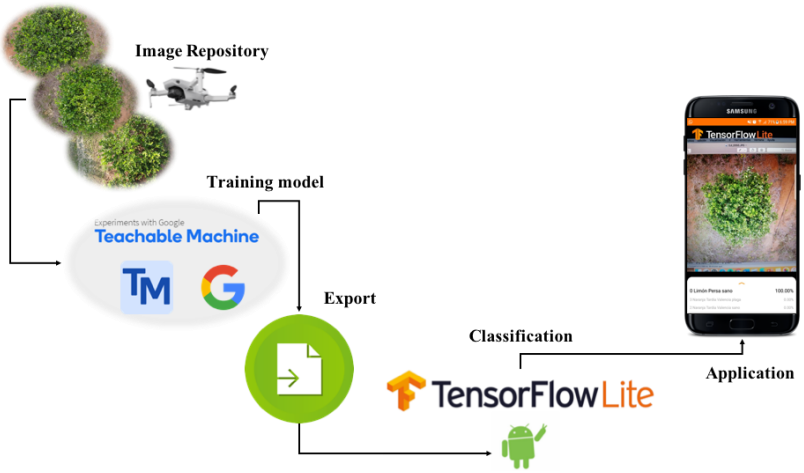


Fig. 9. General description.

The network "input" block comes first. Here, the digital media enters the system (in figure 3 on the top left). Then there is the "learning" block (top, middle), where the model learns and interprets the input. The "output" block (top right), is where the interpretation of the model does something like recognizing the inputs and categorizing the results, so it is important to prepare and load the data set in the teachable machine platform with the Google site as well as define the number of classes, for the case of the project were defined 4 classes. The image classification model was trained there and finally, it was exported.

During the training, you can change the number of hyper-parameters as:

– Number of periods,
– Lot size,
– Learning rate.

Then, it was necessary to export the model in TensorFlow lite format for implementation in Android devices. The ".tflite" files must be placed in the assets

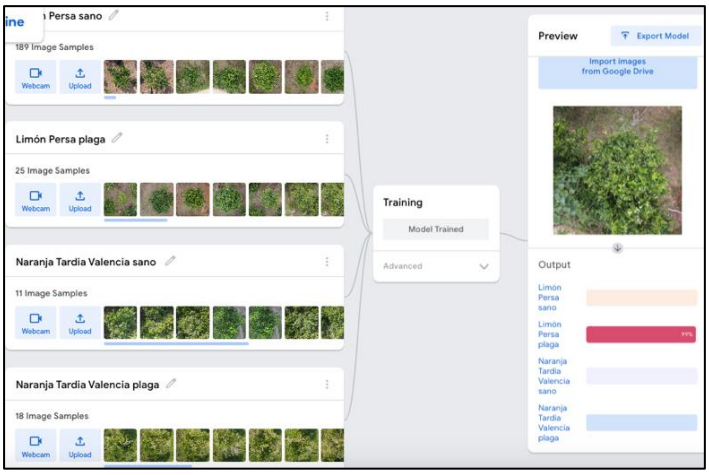


Fig. 10. Training the model.

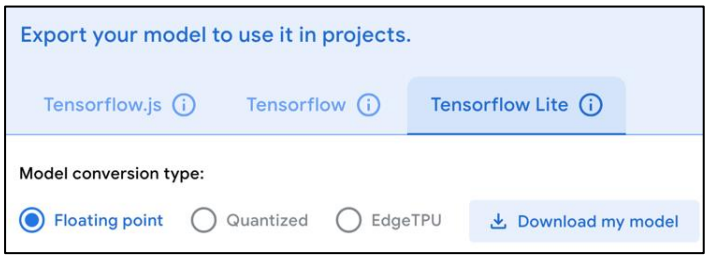


Fig. 11. Exporting the model.

folder of the Android project directory and renamed in the java file that is reading it. In addition, you can download the tflite quantified and FLOAT file format.

2.5 Results

The identification of the type of culture has been obtained by means of the model as well as the classification of the images by means of the convolutional neural network as shown in figure 13. Summarizing: it is possible to say that the important elements that were used to create the CNN were the input layer, this is the pixels of the image, providing height, width and depth working with 3 for Red, Green, Blue.

The Convolution Layer then performs the process of the output of neurons that are connected in "local regions" of input (ie nearby pixels), calculating the product scale between their weights (pixel value) and a small region to which they are connected in the input volume. On the other hand, the Relu layer will apply the activation function on the elements of the matrix and the sampling or subsampling will be in charge of making a reduction in the height and width dimensions, but the depth is maintained, finally the traditional layer of feedforward neuron network that will connect with the last layer of subsampling and will end with the amount of neurons that we want to classify.

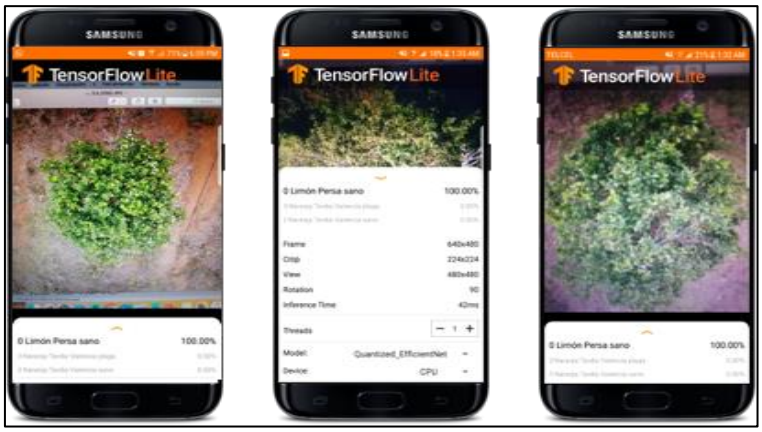


Fig. 12. Image processing.

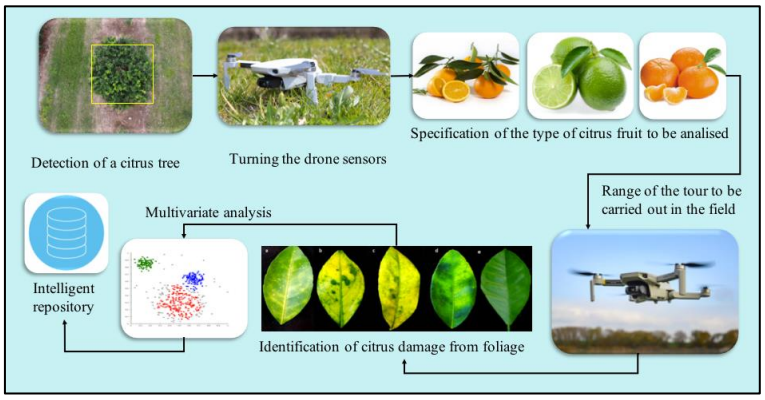


Fig. 13. System layout.

Due to the incorporation of the images and the training of the model it was possible to obtain an identification of the type of culture of 98%, while for the classification of healthy culture and culture with plague the precision in minor, obtaining 88% since at the moment the set of data is unbalanced and this affects the accuracy of the model.

Previously in the figure 10, the operation of the system is appreciated in a general way, starting from the top left, the detection of the type of citrus tree, the sensors of the drone are activated, later it makes the specification of the type of fruit to be analyzed within a range determined by the user, later, it carries out the identification of the damage of the foliage from taking characteristics of color, as following step a multivariate analysis with the purpose of determining the contribution of the factors.

3 Conclusions and Future Research

The original approach of the project has been the study in the fields of deep learning as well as convolutional neural networks. The different methods and techniques used during the development of systems for image recognition have been explored. During

the development several of the original decisions were alternated such as taking into account the recognition of objects within the image. It has been possible to verify the importance of the performance of a neural network since it is not only based on the architecture of the network but the data set available alters the final result, so a balanced data set with a large amount of samples is critical.

The main complications presented are divided into two more relevant ones, image processing and the model. The constructed data set has largely overwhelmed the final accuracy of the application, requiring a much larger data increase than what has been done currently. However, despite providing non-optimal results in some cases, it offers a clear insight into the importance of the available data, as well as the functioning of the neural networks in terms of image perception and how they are treated.

The task of predicting what an image symbolizes is called image classification. An image classification model is trained to recognize various kinds of images. An example of this is the model that has been worked with previously which was trained to recognize images of citrus crops. Thus, when a new image is provided as input to the model, it will generate the probabilities that the image represents each of the crop types it was trained with. In addition, during training, an image classification model receives images and their associated labels, where each label is the name of a different concept, or class, that the model will learn to recognize with enough training data often hundreds or thousands of images per label, an image classification model can learn to predict whether the new images belong to one of the classes in which it has trained to what this prediction process is called inference.

To perform the inference, an image is passed as input to a model. Then, the model will generate a series of probabilities between 0 and 1. Each number in the output corresponds to a label in the training data. By associating the output with the labels with which the model was trained, it was possible to see that the model predicted a high probability that the image represents a pest crop.

The accuracy of the model measured in terms of the frequency with which the model correctly classified an image, which generates a set accuracy of 98%, the performance in terms of the amount of time it takes a model to run inference, which it does in real time since the less time, the faster the model. Also, the disk size varies with its performance and accuracy. The size can be important for mobile development where it can affect the download sizes of the application or when working with hardware where the available storage can be limited, this is the case of the 3.4 Mb model used.

References

1. Bokolonga, E., Hauhana, M., Rollings, N., Aitchison, D., Assaf, M. H., Das, S. R., Biswas, S. N., Groza, V., Petriu, E.M.: A compact multispectral image capture unit for deployment on drones. Conference Record - IEEE Instrumentation and Measurement Technology Conference, pp. 2–7 (2016). doi.org/10.1109/I2MTC.2016.7520445.
2. Briega, R.E.L.: Redes neuronales convolucionales con TensorFlow (2016) https://relopezbriega.github.io/blog/2016/08/02/redes-neuronales-convolucionales-con-tensorflow/.
3. De Rango, F., Potrino, G., Tropea, M., Santamaria, A. F., Palmieri, N.: Simulation, Modeling and Technologies for Drones Coordination Techniques in Precision Agriculture. In Advances in Intelligent Systems and Computing, 873, Springer International Publishing

(2019) doi: 10.1007/978-3-030-01470-4_5.
4. Florida, U.: UF/IFAS Citrus Extension: Plant Pathology (2018) https://crec.ifas.ufl.edu/extension/greening/symptoms.shtml.
5. Gonzalez, T.F.: Handbook of approximation algorithms and metaheuristics. Handbook of Approximation Algorithms and Metaheuristics, pp. 1–1432. (2007) doi: 10.1201/9781420010749.
6. Matich, D.J.: Cátedra: Informática Aplicada a la Ingeniería de Procesos-Orientación I Redes Neuronales: Conceptos Básicos y Aplicaciones (2001)
7. Mead, F.W.: Oficinas Regionales sag-senasa enverdecimiento de los cítricos (Huanglongbing) (2020)
8. Machinelearning: Convolutional Neural Networks: La Teoría explicada en Español | Aprende Machine Learning (2018) https://www.aprendemachinelearning.com/como-funcionan-las-convolutional-neural-networks-vision-por-ordenador/.
9. Pentón Saucedo, Á., del Castillo Serpa, A.: Aplicación de la tabla ortogonal en el diseño de los casos de prueba de software. Avanzada Científica, 15(2), pp. 1–12 (2012)
10. Rivas-Asanza, W., Mazon-Olivo, B., Mejía-Peñafiel, E.: Capítulo 1: Generalidades de las redes neuronales artificiales. Redes Neuronales Artificiales Aplicadas Al Reconocimiento de Patrones, pp. 11–35 (2018) http://repositorio.utmachala.edu.ec/handle/48000/12499.
11. SEDARPA: Veracruz Estadísticas Agrícolas (SEDARPA) (2017) http://www.veracruz.gob.mx/agropecuario/estadisticas-agricolas/.
12. Ulasovich, K.: Teachable Machine, la inteligencia artificial de Google, te responde con un GIF | N+1: artículos científicos, noticias de ciencia, cosmos, gadgets, tecnología (2016) https://nmas1.org/news/2017/10/09/inteligencia-artificial.
13. Vatalaro, F., Mazzenga, F., Giuliano, R.: The Sub-Band Vectoring Technique for Multi-Operator Environments. IEEE Access, 4, pp. 3310–3321 (2016) doi: 10.1109/ACCESS.2016.2580198.

ISSN 1870-4069

Predicting the Survival of Titanic Passengers Using Machine Learning and Smart Data Cleaning

José Alberto Hernández-Aguilar[1], Julio César Ponce Gallegos[2], Víctor Hugo Pacheco Valencia[1], Juan Carlos Bonilla Robles[1]

[1] Universidad Autónoma de Morelos, Mexico

[2] Universidad Autónoma de Aguascalientes, Mexico

jose_hernandez@uaem.mx

Abstract. In this paper we discuss how machine learning can be used to predict the survival of Titanic passengers, we present the general pipeline to predict the survival of passengers, and we focus in the first stage on what we call smart data cleaning, this process can reduce the number of variables and increase precision, recall, and f-score metrics. We compare eight machine learning algorithms: Logistic Regression, Decision Tree, KNN, Gaussian Naïve Bayes, Perceptron, LSCV, Random Forest, and Stochastic Gradient Descend (SGD). The best results were obtained with cross-validation and Logistic Regression with a precision 0.8238, Perceptron 0.8142, and SGD 0.8142.

Keywords: Machine learning, smart data cleaning, survival of passengers.

1 Introduction

An example of a ship disaster is the Titanic, it sunk in the Atlantic sea in April 1912, only (32%) 722 passengers survived of a total 2224 and its crew, titanic sank after hitting an iceberg [1, 2]. While there was some element of luck involved in surviving, it seems some groups of people were more likely to survive than others. According [3] the interesting observation is that in the case of Titanic, some people were more likely to survive than others, like women, children were the ones who got the priority to the rescue, like in Hollywood movies script "children and ladies first".

The purpose of this research is to predict passenger survival rate on the Titanic using different machine learning models and intensive data cleaning. In machine learning, data is divided into Training and Testing, the split ratio could be 70-30 or 80-20; in this research, we use the last one, we have 1047 entries for train and 262 for the test. We use Scikit-learn Machine Learning in Python [4, 5] for making our experiments.

Table 1. Related recent work.

Author(s)	Technique	Results
AM Barhoom, AJ Khalil, BS Abu-Nasser, MM Musleh (2019) [1]	Neuronal networks	99.28% accuracy
B. Balakumar, P. Raviraj, K. Sivaranjani (2019)[7]	Various machine learning algorithms namely Logistic Regression, Naive Bayes, Decision Tree, Random Forest	94.26 accuracy with Logistic Regression
Farag, N., & Hassan, G. (2018, May) [8]	Decision Trees and Naïve Bayes	The Decision Tree algorithm has accurately predicted 90.01% of the survival of passengers, while the Gaussian Naïve Bayes witnessed 92.52% accuracy in prediction
Kakde, Y., & Agrawal, S. (2018)[3]	LR, Decision Tree, Random Forest, and SVM	LR. 0.8372 best result of Accuracy
Tabbakh, A., Rout, J. K., & Rout, M. (2020) [9]	logistic regression, k-nearest neighbors, SVM, naive Bayes, decision tree, and random forest	NA

2 Related Work

2.1 Machine Learning

According [6] insufficient quality of data was the second biggest obstacle to employing AI, narrowly behind a shortage of internal talent. In Table 1 is shown related work of recent research results of predictions about Titanic survivor's prediction.

As shown in the above table, there is a lot of research discussing the prediction of survivors of Titanic, this is a due database of Titanic's passengers is used in several courses of machine learning around the world. The best results were obtained by using neuronal networks reported in [1], and with Logistic Regression in [7].

2.2 Smart Data Cleaning

Quality of data has been analyzed from early days of computing, with the emergent techniques for big data and machine learning techniques this is mandatory to carry out, in [10] is discussed different faces of data quality in the context of this new scenario, in this research tools available and trends are discussed to going beyond just data cleaning.

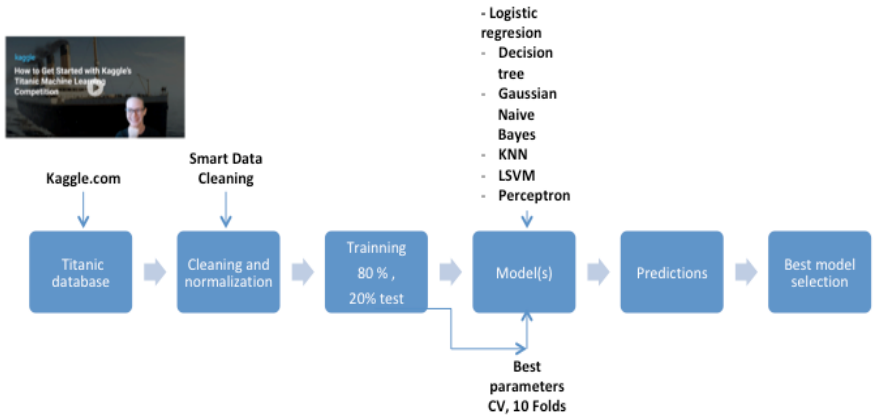


Fig. 1. Proposed Methodology (own source).

Data base	Kaggle	Used
Train	891	1047
Test	418	262
Total	1309	1309
Proportion for test	31.9%	**20%**

Fig. 2. Test and Train proportion Used in this research (own source).

3 Methodology (Pipeline)

The proposed methodology includes six stages: database selection, cleaning and normalization, Training and Test, model selection, prediction, and best model selection (see Fig. 1).

Next, we will discuss briefly each step.

3.1 Database

The database used was obtained from the Kaggle competitions [2] web page, data is publically available. We used 80% of data for training and 20% for testing stages (see Figure 2).

3.2 Data Cleaning

This stage is one of our main contributions, we called it Smart data cleaning, this idea was inspired in data engineering applied by large IT companies like IBM and Oracle, the algorithm is shown below:

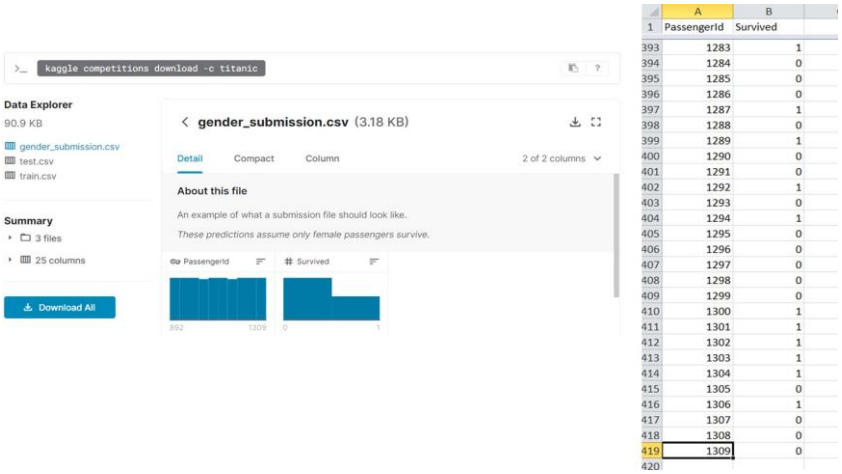


Fig. 3. Detail of gender submission and classification of survivors according to Passenger ID.

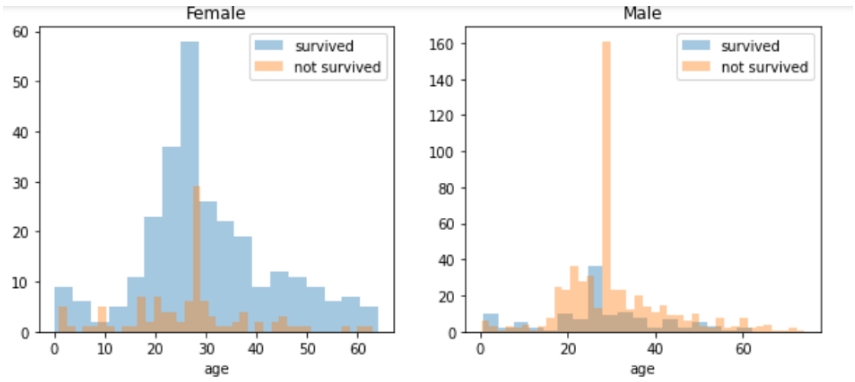


Fig. 4. Detail of gender versus age contrasted survived and not survived.

Smart data cleaning

1. Deep knowledge of data
    a. Visualization of data (Human Analyst)
    b. Crosstab
2. Auto-fill of missing data (by imputation through software-agent)
3. Elimination of not useful columns (Only the columns that contribute to the result are used)

Deep Knowledge of Data. To get deep knowledge of data is necessary to identify what is inside the data. Where does the information come from? How was the information generated? In this stage participates the human analyst/expert to improve the pipeline. In next figure is shown in detail the gender of passengers.

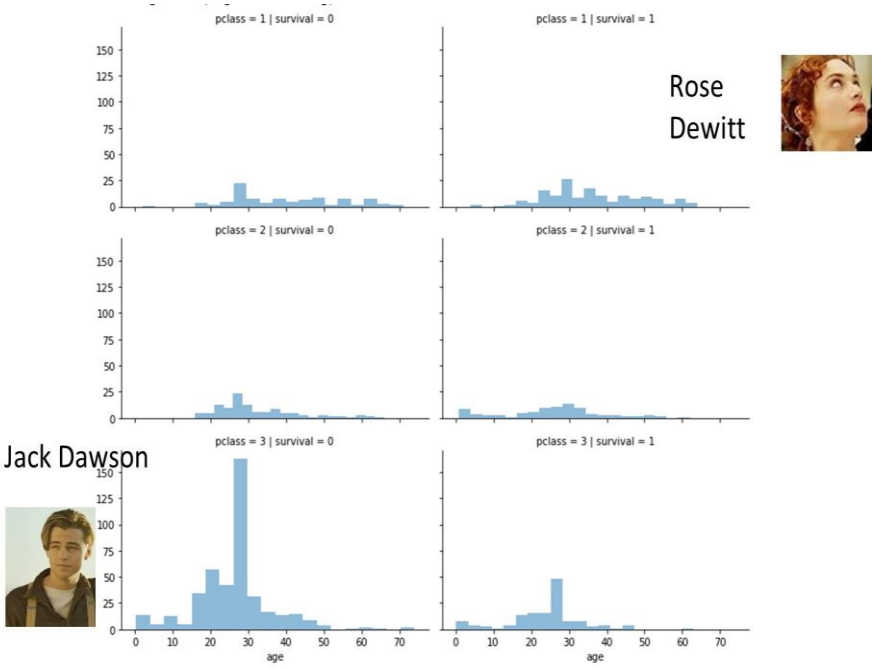


Fig. 5. Social class (pclass) versus age, compared by not survivors (left side) and survivors (right side).

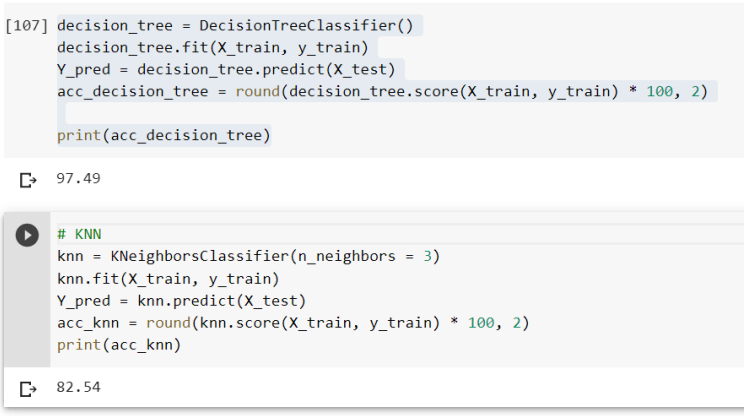
	Total	%
Cabin	687	77.1
Age	177	19.9
Embarked	2	0.2
Fare	0	0.0
Ticket	0	0.0

Fig. 6. Crosstab of fields by missing values.

Data Visualization. To get deep knowledge of data is recommended to visualize data into graphs. The above figure shows what happens during the Titanic's evacuation, the instructions were to abandon the ship "Ladies and children first", therefore there were more females (in all ranges of age) and children who survived. In figure 5, shows how rich first-class women and children survived.

Crosstab. To analyze the quality of data a crosstab of fields will reveal which data must be completed, imputed, or deleted. Figure 6 shows some fields are complete (i.e. Fare Ticket), but some others must be imputed (i.e. Age) or deleted (i.e. Cabin).

As shown in figure 6, there are fields with missing values, but we can improve some of them using auto-filling, as described in the next pseudo-code.



```
[107] decision_tree = DecisionTreeClassifier()
      decision_tree.fit(X_train, y_train)
      Y_pred = decision_tree.predict(X_test)
      acc_decision_tree = round(decision_tree.score(X_train, y_train) * 100, 2)

      print(acc_decision_tree)

      97.49

      # KNN
      knn = KNeighborsClassifier(n_neighbors = 3)
      knn.fit(X_train, y_train)
      Y_pred = knn.predict(X_test)
      acc_knn = round(knn.score(X_train, y_train) * 100, 2)
      print(acc_knn)

      82.54
```

Fig. 7. Details of Decision tree and KNN models.

Auto filling of missing data

1. If columns are not promising (i.e. Too empty or do not contribute)

   If (ttnc_df[i] >= 70%)

   Delete column i

   If (ttnc_df[i] is not correlated)

   Delete column i

2. If data is incomplete but could be useful

   Auto-fill column by imputation (i.e. Median value, or random values)

3. Convert text fields to categories (numerical data)

3.3 Training and Test

We use 80% of data for training and 20% for testing.

3.4 Model

We train and test the next models: Logistic Regression, Decision Tree, KNN, Gaussian Naïve Bayes, Perceptron, LSCV, Random Forest, and Stochastic Gradient Descend (SGD). In further research, we will describe each of these models.

3.5 Predictions and Model Selection

We obtain predictions for each model, first without cross-validation, and later with 10 folds cross-validation. For our experiments, we used Google Collab, Python 3, and set up the environment GPU enabled. The results are shown in the next section.

Table 2. Results of prediction.

	LR	**Deci-sion Tree**	**KNN**	**Gaussian Naive Bayes**	**Perceptron**	**LSVC**	**Random Forest**	**Stochastic Gradient Descent (SGD)**
Proposed pipeline								
Precision	0.7859	**0.9749**	0.8254	0.7691	0.6938	0.7656	**0.9749**	0.7392
Donges (2018)								
Precision	0.8114	0.9282	0.8732	0.7710	0.8070	0.8081	0.9282	0.7699
Opt. w/ CV	**0.8238**	0.7619	NA	NA	**0.8142**	0.8333	0.7714-**0.8142**	**0.8142**

4 Results and Discussion

4.1 Predictions

Table 2 shows the results of predictions, the first row shows the precision of the proposed pipeline, the second row shows precision obtained in [11], the third row shows the results of our proposal using ten folds cross-validation (CV).

The main results are shown in black, as can be seen, the best results are obtained with the proposed pipeline and Logistic Regression with ten folds cross-validation. Perceptron and SGD get second-best results 0.8142; in these three cases, precision was better than results obtained by [8].

ROC AUC CURVES of [8] and of our most representative results are shown in the next figures. A classifier that is 100% correct, would have a ROC AUC Score of 1 and a completely random classifier would have a score of 0.5.

As can be seen in Table 2, and figures 9 and 10; our results were better than those reported by [11]. Our best result of Accuracy (0.8238) was obtained with Logistic Regression which is promising according to results obtained in [7] and [3].

5 Conclusions and Future Work

The smart data-cleaning algorithm employed as a stage of machine learning demonstrates its functionality to improve the results of prediction for survivors of titanic.

Our future work will be to apply the proposed methodology in different databases, including synthetic and real-world databases, for instance in bank databases to analyze information from users to predict payment of credits. In the case of applications in education, we plan to predict performance in online assessments based on student habits, previous grades, preferences, and usage of LMS during this period of the pandemic.

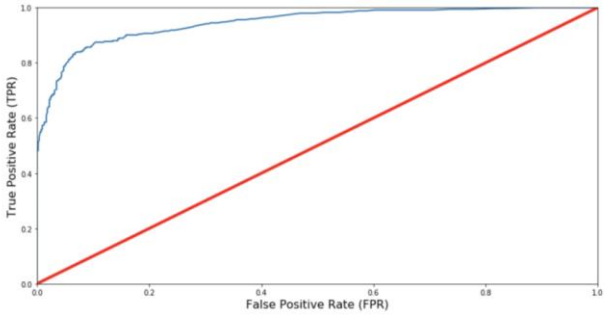


Fig. 8. Random Forest Classifier, precision 0.8019, Roc AUC Score=0.9450 [11].

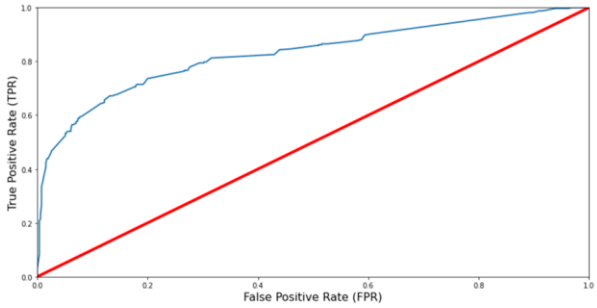
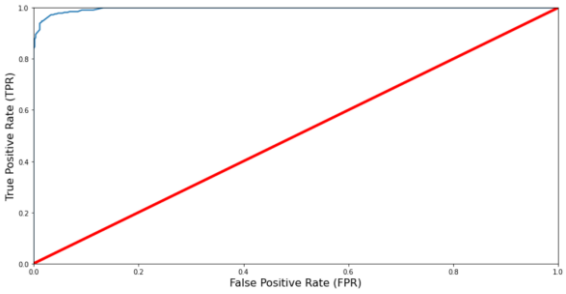


Fig. 9. Logistic Regression, precision 0.8238, ROC AUC Score = 0.8286.



.

Fig. 10. Random Forest, precision 0.8142, ROC-AUC-score = 0.9963.

References

1. Barhoom, A.M., Khalil, A.J., Abu-Nasser, B.S., Musleh, M.M., Abu-Naser, S.S.: Predicting Titanic Survivors using Artificial Neural Network. Int. J. Acad. Eng. Res., 3, pp. 8–12 (2019)
2. Kaggle: Kaggle (2019) https://www.kaggle.com/c/titanic.
3. Shefali Agrawal, I.: Kakde Asst Professor AITR, Predicting Survival on Titanic by Applying Exploratory Data Analytics and Machine Learning Techniques.

Artic. Int. J. Comput. Appl., 179, pp. 975–8887 (2018) doi: 10.5120/ijca2018917094.

4. Scikit-learn: Scikit-learn: machine learning in Python — scikit-learn 0.23.1 documentation (2020) https://scikit-learn.org/stable/.
5. Raschka, S., Mirjalili, V.: Python Machine Learning: Machine Learning and Deep Learning with Python (2020) https://books.google.com.mx/books?hl=es&lr=&id=sKXIDwAAQBAJ&oi=fnd&pg=PP1&dq=Raschka,+S.+and+Mirjalili,+V.+(2019).+Python+machine+learning&ots=V8LmsSTDEn&sig=Zxkoa3nncSs4gJvLlg8IUlO9DAA&redir_esc=y#v=onepage&q=Raschka%2C S. and Mirjalili%2CV.
6. MIT, T.R.I.: Insights survey: Talent shortage is the top AI challenge – MIT Technology Review Insights (2020) https://insights.techreview.com/live-ai-poll-key-stats/.
7. Balakumar, B., Raviraj, P., Sivaranjani, K.: Prediction of survivors in Titanic dataset: A comparative study using machine learning algorithms (2020) https://pdfs.semanticscholar.org/545a/9e5da57058cf08e32eae6b5816839505ac3c.pdf.
8. Farag, N., Hassan, G.: Predicting the Survivors of the Titanic-Kaggle, Machine Learning From Disaster (2018) doi: 10.1145/3220267.3220282.
9. Tabbakh, A., Rout, J.K., Rout, M.: Analysis and Prediction of the Survival of Titanic Passengers Using Machine Learning. In: Lecture Notes in Networks and Systems, pp. 297–304, Springer (2021) doi: 10.1007/978-981-15-4218-3_29.
10. Gudivada, V.N., Ding, J., Apon, A.: Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations Big Data Management View project Cognitive Computing (2017)
11. Donges, N.: Predicting the Survival of Titanic Passengers (2020) https://towardsdatascience.com/predicting-the-survival-of-titanic-passengers-30870ccc7e8.

ISSN 1870-4069

Zinc Recycling in Die Casting Processes Using Industrial Optimization

José Guadalupe Velázquez Ramírez, Alberto Ochoa-Zezzatti

Universidad Autónoma de Ciudad Juárez,
Mexico

Al194549@alumnos.uacj.mx

Abstract. Zinc is one of the metals existing on the planet too important for the life of this one, it is an element of the essential nature in our development of life of every living microorganism. Agriculture is one of the productive processes of the planet that undoubtedly depends a lot on the quantities of zinc that exist in nature. The pharmaceutical industry has generated thousands of products based on this element to generate medicines for human health. But we must not forget that despite being a necessary element it is also toxic for living beings, and if we wanted to eliminate it from our environment, we would also cause an immense problem to any ecosystem or to our own environment. The excessive use of metals has generated an excessive pollution to the planet, and what is required is to create awareness in the industry to generate a green market. It is required that all companies begin to think about giving back some of what they have taken from the planet, making improvements in their processes to create products with sustainability to have a world without pollution. Technological research has developed new improvements to many applications within the industrial processes that have placed many companies in the first places as recycling plants and generating products based on zinc waste and remain as green companies according to the rules. Generating an evolution to the economy of the main producing and consuming countries of this element, which in spite of being less required than aluminum, in the last years, from 2000 to 20018, more than 45700 million metric tons have been generated all over the world. China is the country with the highest zinc productivity in the world.

Keywords: Die casting, Zamak zinc alloy, zinc recycling.

1 Introduction

In this article we analyze the importance that zinc has in the natural, environmental, after its normal and industrial process cycles as a recycled material. The aspects caused by this type of activity that lead to an increase in the economy in the countries that are dedicated to the collection of this metal are analyzed. Nature on the planet has evolved with the presence of zinc, playing a fundamental role in several biological processes and being an essential element for all forms of life. Small microorganisms, living beings, including human beings, without this material their existence would not be possible.

- Zinc is a natural element, for this reason it cannot be eliminated from our environment, if this happens our ecosystems could have very harmful effects in the long term. This element is indispensable for living beings and it is necessary for humans to ingest it in their daily diet in small doses. It is necessary to know that this element, despite being an important part of life, also becomes toxic when consumed in excessive quantities (Miller, 2002).
- The intoxication can occur by different ways: inhalation, by consumption of drinks or by ingestion. Intakes of 50-60 mg of zinc/day can cause abdominal pain, nausea, vomiting and occasionally pancreatitis, especially if accompanied by inadequate copper intake (Sandstead HH, 2006). Environmental toxicity also exists from elevated levels of zinc released by various human activities.
- Emissions come from municipal and industrial waste, mining, geological activities, and secondary sources not related to zinc production or manufacturing. Current uses of zinc and zinc compounds alone do not lead to the high levels found in some regions in water and sediments. The bioavailability of zinc transport in water, sediment, and soil depends on the chemical characteristics of the environment and organisms on the site (Tejeda S., 2005).

2 Normal Zinc Cycle

Zinc is a common mineral on the planet located in soil, rocks, water and air. Natural emissions to the atmosphere amount to 5.9 million metric tons annually. Sixty percent of the zinc produced worldwide is extracted from mines and the remaining 40% comes from secondary process waste. However, recycling is increasing year by year, with progress being made in zinc production technologies IZA (International, 2014).

- Zinc follows a complex life cycle from extraction as a mineral, through refining and use in society. This life cycle can be characterized by the collection of information at various stages of production, manufacturing, use and waste management to the eventual collection and recycling of products.
- Zinc is an extremely versatile material, which plays a fundamental role in certain industrial applications and products that generate lower costs and a lower environmental impact due to reduced maintenance. In the end, the material recovered from modern products that are discarded can be recycled without deterioration (Martinez J. Roca, 2001).
- Zinc during its normal application is sustainable as it is recyclable at all times of use, playing an important role during the final life phase by consuming little energy, reducing emissions and minimizing waste disposal. Two commonly used approaches to evaluate the percentages of sustainability are the amount contained in the material processed, and the fraction of zinc obtained at the end of its life.
- Zinc is very useful in the chemical and metallurgical industry because its physical and chemical characteristics are very useful for products made from this material. It is a material that can be melted at very low melting point temperatures, 420°C and boiling point of 907°C (SANCHO J., 1999).

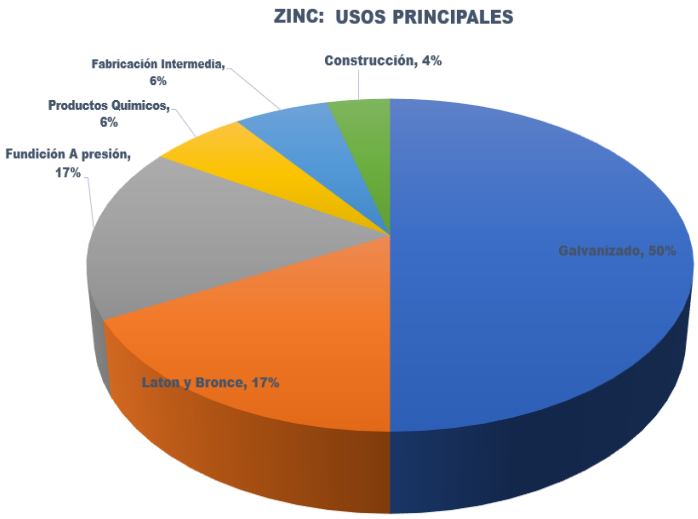


Fig. 1. Zinc as the main metal in the industry (IZA(International, 2014).

– This material is fragile at room temperature, but after 100°C it becomes malleable. Its resistance is higher than that of lead and tin, but considerably lower than that of aluminum, and despite its low nobility it has very good resistance to atmospheric corrosion.

3 Main Uses of Zinc

The International Zinc Association (IZA), in cooperation with Yale University, completed a flow analysis of anthropogenic zinc emissions into the atmosphere based on zinc production and use figures. These analyses and studies have been supported by the industry to understand the extent of environmental impacts beyond the life of zinc IZA (International, 2014).

In this study, the figures produced annually around the world by the transformation industry are presented, which were 11 million tons of zinc use. Of these, the recycling rate at the end of its useful life was 50%, the amount used.

Of the brass and bronze alloy 17% is used for production mainly in the die casting industry. Significant amounts of 4 and 6% are also used in compounds, for example, zinc oxide, zinc sulfate, for the manufacture of sheets required in construction placed on roofs, gutters and downpipes of houses or buildings.

4 Zinc Nonferrous Metal

Zinc is one of the main non-ferrous metals with atomic number 30, an atomic weight of 65.38 and a specific weight of 7.14g. It is a grayish white metal resistant to oxidation and corrosion; once melted it has a high fluidity with a tensile strength of 2kg/mm2,

increasing when processing its lamination. Zinc has a high elasticity in hot of 120-150°C, keeping its malleability, but being harder than tin.

This material is used in manufacturing along with Aluminum, Copper, Magnesium, Nickel, Lead and Titanium. Non-ferrous metals are used in manufacturing processes, as complementary elements to ferrous metals, are also very useful as pure or alloyed materials, which by their physical properties and engineering cover certain requirements or working conditions, for example, bronze (copper, lead, tin) and brass (copper zinc).

Being a non-ferrous metal, it becomes one of the most important natural qualities for use in the coating of steels to prevent corrosion in this type of ferrous material. Steel in any environment if left unprotected corrodes very easily and quickly, hence the use of zinc as a physical and protective barrier (Krúger J., 1990).

Pure steel in its galvanizing process undergoes a caustic cleaning and annealing before the molten zinc coating is applied. A zinc chloride and ammonium flux is used to remove the last traces of oxide. This process must be applied at a speed of 200 meters per minute to avoid the cooling of both metals causing the zinc to not adhere properly to the steel.

Continuously galvanized coatings metallurgically adhere to the steel they protect. This ensures the adhesion of the coating, which is essential for manufacturing processes that bend, stamp, roll or draw the steel into its final product form.

5 Zinc Alloys Used in Die Casting

During the 1930s, many of the alloys we know today were available, but their use was very limited. Modern science and technology, metallurgical controls and research are now making refinements possible that result in new alloys with greater strength and stability.

Over time this research into different alloys has led to significant technological improvements in the basic process of high-pressure zinc injection molding applications (die casting). These alloy improvements have been effective in expanding die casting applications in all known markets.

Among the technological investigations is zamak, which belongs to the 17% zinc-based alloy used in the die casting industry, which is composed of various non-ferrous metals to give the necessary purity to the alloy. Zamak is manufactured in six types and several sub-types and its basic composition contains besides Aluminum, Magnesium and variable amounts of Copper and Nickel, these metals are added with the purpose of improving the mechanical properties of the base metal, in this case zinc.

It is very important for the die casting process that the high-pressure molding designers, before making any application project should consult a table containing the mechanical properties of Zamac alloys. These properties of typical interest are:

- Tensile strength (maximum).
- Production strength.
- Elongation (ductility).
- Modulus of elasticity (MOE).

Fig. 2. Zinc Alloy Ingot (SoloStocks, 2020).

Table 1. Zamak Alloys (NADCA, 2015).

Tipo de Aleación	Aleación de Zamak				Aleación ZA		
	#2	**#3**	**#5**	**#7**	**ZA-8**	**ZA-12**	**ZA-27**
Nominal	Al-4.0	Al-4.0	Al-4.0	Al-4.0	Al-8.4	Al-11.0	Al-27.0
Composición	Mg-0.035	Mg-0.035	Mg-0.055	Mg-0.013	Mg-0.023	Mg-0.023	Mg-0.015
	Cu-3.0		Cu-1.0	Cu-0.013	Cu-1.0	Cu-0.88	Cu-2.25

Each of these is a property which predicts how the alloy will react to a stress condition. A "strong" alloy has high tensile and elastic strength values and low elongation values. A "weak" alloy has low strengths and higher elongation values.

Zinc alloys are referred to as Zamak, which is an acronym for “Z” for zinc, for “a “aluminum, “m” for magnesium and “k” for copper. In the United Kingdom, they are called Mazak Zamak is manufactured only with "Special High Grade" Zinc, with 99.99+% purity, pure Aluminum and Magnesium and electrolytic Copper. Its most common presentation is in ingots of 8 kilograms approximately, the standard stowage handles 132 ingots (although it can be adjusted according to needs) and weighs ± 1100 kgs.

Existing Zamak ingots comply with the following standards: NOM-W-137-85 and ASTM B240-87 and they may contain Nickel, Chrome, Silicon and Manganese in quantities that correspond to their solubility of 0.02%; 0.03%; 0.035%; and 0.5% approximately and respectively at freezing temperature.

The following table shows the elementary specifications for the zinc alloys normally used in die casting. Zamak #3 and #5 were introduced in the 1930's and #7 were introduced about 20 years later.

In this type of productive activity, waste of this metal is generated in three stages that are used to create a good that will generate a production cost. During the processes of melting the zinc ingots in the production equipment, there are several processes of

consumption of this metal in which a percentage of the purity of the virgin Zamak is constantly lost.

The zinc based Zamak alloys, as mentioned above, applied in the industry are 17%, zinc die casting scrap, zinc dust filtered by the steel mill furnaces represents 6% as well as steel plates.

6 Recycling in Die Casting

Slag

The slag is basically formed from the material coming from the furnace or crucible, generated by the high temperature in the melting process, this temperature is normally 800° Fahrenheit and is recovered in an impure form of 50% zinc. This material is a kind of foam that is generated on the surface of the molten metal inside the container (Crucible). Its chemical composition is basically zinc and iron.

Flash

The burr is a material that is generated at the time of molding the main part that is already the finished product. This is a surplus material or an excess that is formed to the figure of the piece and affects the quality of your design, so all parts that come out with this defect are recycled as virgin zinc. The material that is discarded during the injection process in a die casting process does not lose the property of being a pure zinc alloy. This material is still Zamak and for this reason it is returned to the production areas and placed back in the molten zinc container furnaces (crucible) to be recycled during the same process. The special dies used to generate the die castings are provided with several slots that serve to dislodge the injected zinc gases into the die cavities. These grooves generate several excesses of zinc that are not used in the main product, so they are also recycled, as is the burr mentioned above.

Injected material

The amount of leftover material that is used at the time of injection in the mold dedicated to the molding of specific parts, not all this material is used and this is returned to the furnace (crucible) and this zinc waste is also collected by the company and sent for recycling.

Material with Quality Defects

Another very important waste in this recycling activity is the pieces that have a final finish, this means that they are pieces that after being manufactured go through other processes required by the customer's design, such as machining, polishing and sanding. During these processes, the pieces suffer damage and are rejected by the quality area and become useless material. There are already finished materials that, take secondary processes such as painting, which is a finish required by the customer.

This material commonly comes out with defects in the quality of the paint, 'so it is rejected and sent to companies dedicated to the recycling of this material to give them their proper process and generate the new alloy with the properties of Zamak again.

The parts already finished and available as scrap, like all the burr and dust generated in secondary processes, which are the machined, these can be placed directly in the ovens (crucibles) as this material still has all the characteristics of a completely virgin zinc alloy.

Companies that perform these zinc-based processes are committed to the principle of sustainability, and this commitment is rooted in the sustainability charters and guiding principles. The companies believe that protecting the environment, openly engaging in sustainability issues, and supporting sustainable development practices not only drives long-term prosperity for the industry and enables customers to become more sustainable through the use of zinc products.

The processing industry knows that nature is the main medium that exists as a supplier of raw materials for the production of the different processes in the industry, and for this reason the consumption of these materials requires their use in the productive areas.

These activities generate waste that, sooner or later, will return to the natural environment and depend on the handling of these, when they are discarded as they normally produce pollution.

It is necessary to know that, the productive sector will reuse some waste in the form of recyclable products, and that there are other materials from production that will not be recycled being thrown into the environment, so that it will behave in its role as a sink when receiving the waste (Martinez, 2013).

The environmental aspect and sustainable development have had a great influence on the metal recycling industry, without forgetting that this has been important in generating growth in the economy of countries. Due to the exploitation of natural resources and the scarcity of exploitable mineral reserves, the recovery of zinc from industrial waste is very important because zinc values can be obtained at very low costs. This increase is due to the fact that the recycling of metal waste is a truly lucrative activity for the world economy (Martinez J. Roca, 2001).

In the U.S. since 1970 30,000 jobs were created marking the economic and ecological importance of recycling until 1985. During this time, the American society annually discards around 35 million aluminum cans, generating $200 million dollars due to the collection of this metal waste, which is indispensable for society's consumption (Sergio Augusto Fernández Henao, 2010).

7 Main Zinc Producing Countries

A growing production of zinc in recent decades was due to its increased demand, especially in the construction sector, given the economic inclusion of China in India. During 2000 to 2018 zinc production increased by 53% to 13.7 million metric tons worldwide (GERENS, 2020).

In terms of world zinc production, the leading producer of this metal in 2018 was China, followed by Peru, as shown in the graph. At regional level, and during this year, the region of Ancash showed the highest production, with 36.8%, followed by the areas of Ju-min and Pasco which were in second and third position. The three regions together represented 71.5% of the national zinc production.

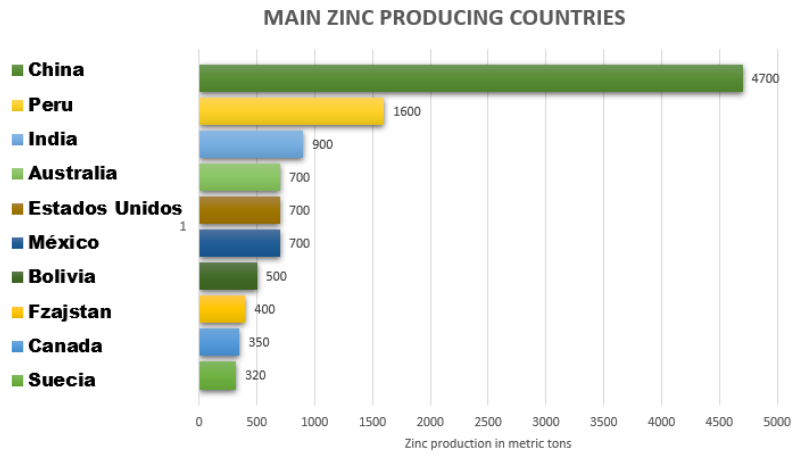


Fig. 3. Main countries in zinc production (GERENS, 2020).

Peru is the second largest producer of zinc in the world and in 2018 the companies Antamina, Volcán, Nexa and Nexa el Porvenir. They were responsible for 62% of the year's production. This resulted in zinc exports worth more than U.S. $2,563 million, representing 9% of Peru's mining production (ANTAMINA, 2020).

At present, zinc is one of the most widely produced industrial metals and is only surpassed by iron, aluminum, and copper, with China, Australia and Peru being the main zinc producers. China is the largest producer of zinc in the world, generating 4,57000 tons of metal used to galvanize steel.

Even though in recent years there have been reductions of 10% due to low prices of this metal, companies in the Asian giant have managed to increase this price by 5.3% in the sale of this non-ferrous metal. Complexity is a characteristic in the process of recycling metals, but the main objective of this activity is the enormous savings in production costs.

This is reflected in the 95% of the little energy that is used to recycle aluminum and zinc, different from that used to create them again. Over time, the increase in new technologies will allow the process to be fast, efficient and with renewable energy. Products that are not only recyclable, but also produced with a maximum of renewable substances and energy, are certified as Cradle-to-Cradle, (C2C) (Business, 2020), (IZA International, 2014).

The concept of Cradle to Cradle aims to close the life cycle of every product, making the end of life a beginning of the next production. This type of economy has created one of the greatest barriers to sustainability due to the current linear economic model of "take-manufacture-dispose" (Business, 2020).

This is done using the model of a circular economy, first used by (Pearce and Turner 1990), which aims to redefine growth by focusing on positive benefits for all of society. New markets are experiencing a rapid change and boom in eco-design, which helps companies stand out from the competition. Today there is a new generation of entrepreneurs concerned about environmental issues that focus their processes thinking

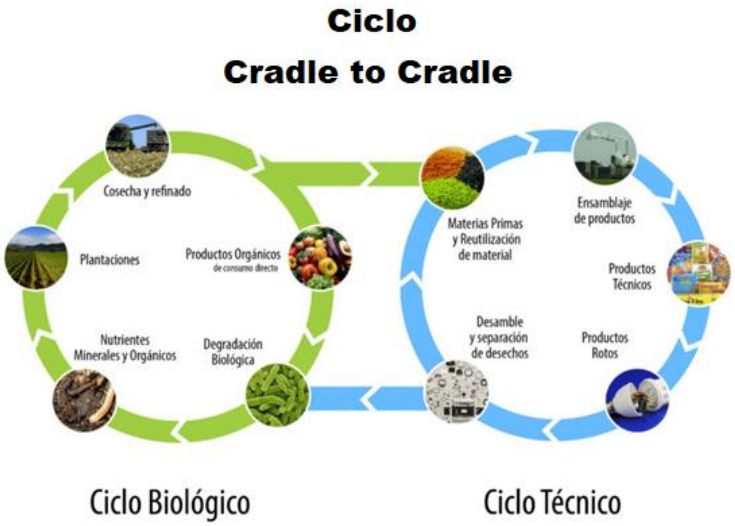


Fig. 4. Life cycle of a product (Miller, 2002).

about the entire life cycle of the product, from the manufacturing process, until the end of the life of a product (Martinez J. Roca, 2001).

These markets need companies to be sustainable and to show the world the need to build non-polluting factories that produce green products called eco-design, safe for the environment and recyclable, developing new industrial methods and analyzing all the raw materials used for manufacturing (Business, 2020).

It is necessary that the products manufactured under this new process have to go with a new brand, which will certify that the company that is elaborating it cares about what happens with this product since it is created until it is discarded by the final user.

References

1. Business, I.C.: Blog de energías renovables (2020) Retrieved from https://blogs.imf-formacion.com/blog/energias-renovables/articulos/mas-alla-del-reciclaje-concepto-cradle-to-cradle/.
2. Doehler, H.: Die Casting. New York: McGraw Hill Book Company (1951)
3. Institute, C.t.: Products Innovation Institute (2020) https://www.c2ccertified.org/get-certified/product-certification.
4. IZA: Zinc un material sostenible (2014) https://www.zinc.org.
5. Krúger, J.: Reciclaje de metales no ferrosos. In Reciclaje de metales no ferrosos, Aquisgrán: Instituto de Metalurgia Extractiva (1990)
6. Martinez, J., Roca, J.: Economía ecología y politica ambiental. In J. Martinez J.Roca, Economía ecología y politica ambiental. México: Fondo de cultura Económica (2001)
7. Martínez, M.B.: El suelo como sumidero (2013)
8. Miller, T.: Ciencia Ambiental Preservemos la Tierra. México: Editorial Thompson (2002)
9. NADCA: Product-Standards-for-Die-Casting. pp. 272 (2015)
10. Sancho J.: Procesos de obtención. In Metalurgia Extractiva, Sintesis S.A (1999)

11. Sandstead H.H.: Zinc Requirements and the resk and benefits of zine supplementation. In J. Trace, Elemen Med Biol., pp. 2–18 (2006)
12. Fernández, S.A., Henao, L.A.: Scientia Et Technica (2010) https://www.redalyc.org/articulo.oa?id=84920977045.
13. Tejeda, S.: La medición de Sedimentos en México ED IMTA (2005)

ISSN 1870-4069

KANAN: An Intelligent Application to Determine Road Accidents Associated with Mental Fatigue in Queretaro within the Paradigm of a Smart City

Aaron Zarate[1], Alberto Ochoa-Zezzatti[2], Luis Maldonado[1]

[1] Universidad Autónoma de Queretaro,
Mexico

[2] Universidad Autónoma de Ciudad Juarez,
Mexico

alberto.ochoa@uacj.mx

Abstract. Queretaro is a large city located along in center of Mexico with a population over a million people in 87 km^2 has recently experienced a situation of violence and insecurity related directly to organized crime: assaults, kidnappings, multi-homicides, burglary, between others. However, the second leading cause of death in the city is associated with traffic accidents: 1,377 deaths in 2019 alone. For this reason, citizens have actively pursued specific programs that would decrease the overwhelming statistics: 3,897 deaths from 2014 to 2018. The reason of the following project is to provide drivers with a technological tool with indicators and sufficient information based off statistics compiled by the *Queretaro Security Government Department* and other public sources. Then drivers would have more information on possible traffic accidents before they happened. This research tries to combine a Mobile Device based on Data Mining and a Kriging Model to determine the danger of suffering a traffic accident associated with mental fatigue of bus drivers in a given part of the city during a specific time. KANAN, means vigilant guardian in the Mayan language is therefore the acronym of our intelligent application of ubiquitous computing

Keywords: Data mining, kriging model and mobile devices.

1. Introduction

According to literature is feasible to use an application to determine the possibility that a traffic accident will occur using data associated with most frequented areas. Then the users will be aware of alternate routes to their destination, among other advantages. Using technological tools, data analysis and mathematical models would then decrease drivers' chances of suffering a traffic accident. The aim of this project is to create a model that uses these tools in a mobile geographic information system (SIGMA) on levels of traffic accidents in specific areas of Queretaro, which is presented in detail below, this research include the analysis of Data Mining to determine the danger that a traffic accident would occur during a specific time. We realize an exhaustive analysis

of other similar research, the only similar context is explain in [6], where the authors calculated the insecurity of a vehicular group which requires product delivery in different places with random scheduling, but this research does not considers real on-time statistics and the perspective to suggest a different route to travel or stay for a determinate period of time.

Kanan is a project that was developed as a proposal to improve the road safety by providing accurate, useful and representative data to understand the factors that creates the conditions for a car accident to occur. The tool, that consists in three different stages is the result of an analysis and investigation performed as a part of a Design and Innovation Master´s degree program from the Universidad Autónoma de Queretaro, in Mexico.

The objective of this experiment is to identify how does the proposal works. Through this experiment, a beta app was used to have volunteer users creating reports of the incidents they observed during their daily activities. The reports coming from the app provided the information that built the database that at the end should allow to identify conflict zones through a geostatistical analysis. In order to rate the project as successful, the system had to provide a clear identification of those conflict areas from now on called red spots.

2. Implementation of the Intelligent Application

There are several aspects that need to be taken into account when designing mobile applications: limited screen size, different resolution and screen sizes across devices. Therefore, the designer has to develop the interface uniformly so that it suits most devices. This module explains how to work with different layouts provided for the Android API. The programming interface we will work on is XML.

There are several ways to obtain the geographical position of the device, however we would thoroughly use GPS and access points (Wi-Fi) nearby; both perform similarly but differ in accuracy, speed and resource consumption. Data Server Communication is the most important module because it allows communication with the server, allowing you to send the GPS position obtained by receiving the processed image and map of our location, thus showing the outcome of your application that is the indicator of insecurity.

3. Test Development

The beta version from the Kanan app was developed with the support of the company Next level, based on the minimum requirements which were the delivery of a very basic app with the capability to interact the user using a simple but practical interface and automatically register within the report generation the date, hour and location given by the mobile device GPS.

The app has in its main page the option for a registration or a login. A registration of the user allows the identification of the user’s profile and demographic information.

Fig. 1. Intelligent Application developed to determine incidence levels of traffic accidents in the Queretaro Metropolitan area.

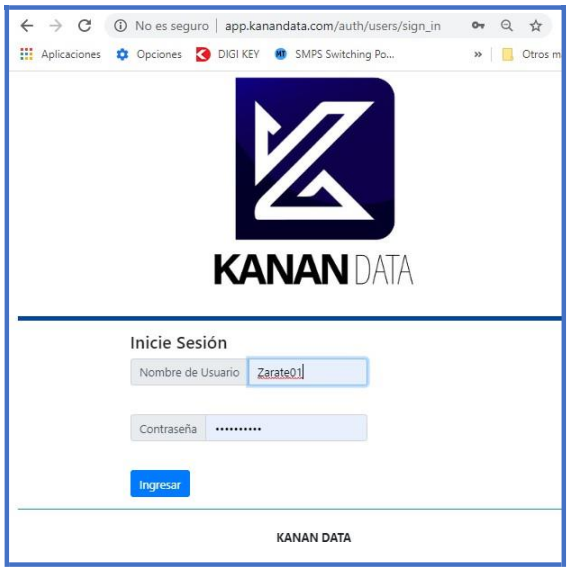


Fig. 2. Connection to a remote server with real data of transport and mobility in Queretaro.

Using our intelligent application (see figure 1) we can determine in detail the prevalence of any traffic accident associated with mental fatigue of drivers in the Queretaro Metropolitan area.

The generated reports were sent automatically to the web-based database, with the domain app.kanandata.com. The reports are listed and classified according to the way they were reported. In order to protect the information, the database requires secured access. In figure 2, we receive information associates with connecting with a specific server which obtain information from Queretaro transport department.

In order to do the test, it was published on social network sites the call to participate in this experiment, therefore, all the 13 registered users were volunteers.

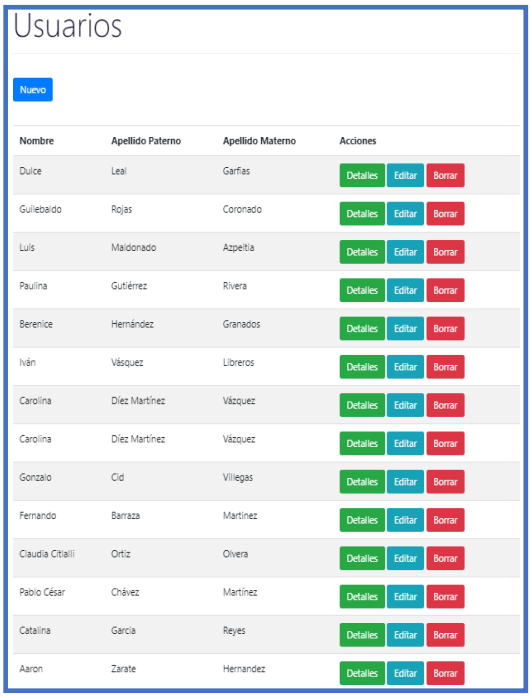


Fig. 3. List of users of our Intelligent application using real data to take a correct decision related with traffic in a Smart City.

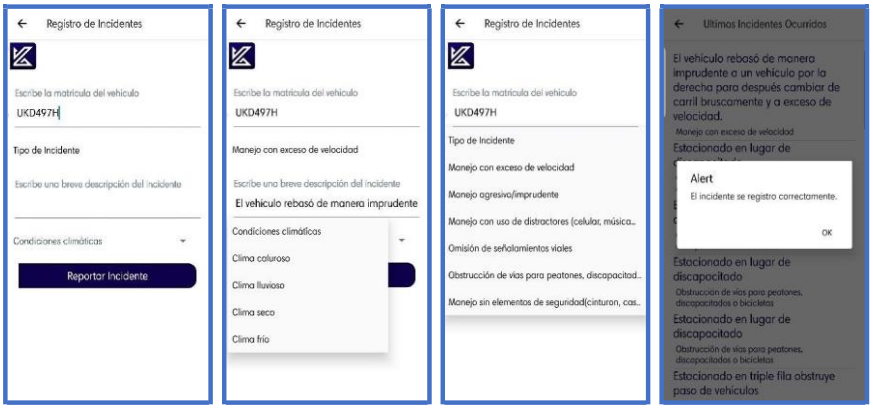


Fig. 4. Textual information related with a new traffic accident involves bus drivers.

Due to the simplicity of the app at this point, where it is still not available a voice interaction feature, which will be intended to avoid the driving users to incur in the usage of the app as a distractor, the selected profile were those people that do not typically drive during their daily journeys, like public transport or bike users, or walkers. In that case, they could stop for a moment to register an incident without putting themselves in risk. Each user obtains new data related with section of city where

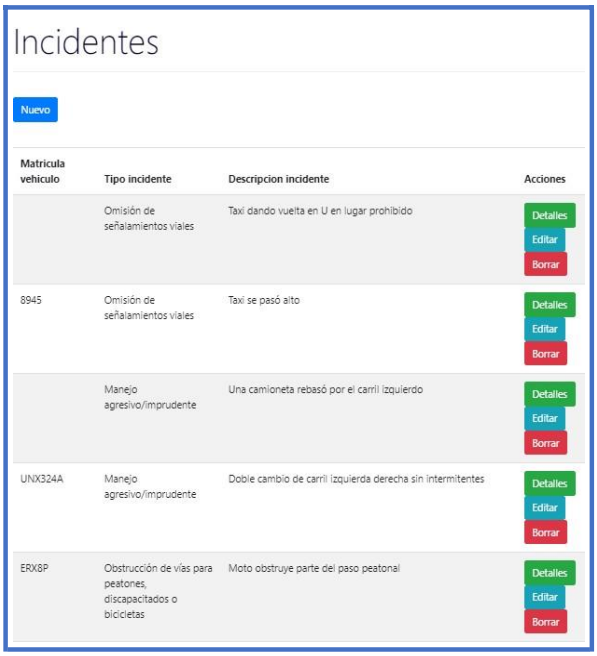


Fig. 5. Relation of each traffic accident by type and geospatial data to built clusters in a Smart City.

occurs an extra traffic to prevent risk of future accidents as is possible see in figure 3.

Using the app, the users could register different information, such as the car plates (if they were able to catch it), the type of incident, a brief explanation of the incident (if feasible) and the weather conditions. Regarding the type of incidents, there is a pop-up menu to select between the 6 different listed types: Speeding; rash driving; driving with distractors; road markings omission; pedestrian or disabled path obstruction; and driving without safety elements (seatbelt, helmet, etc.). As mentioned before, the app register by itself the time, date and location at the moment of the report generation, with no chance for the user to edit this information. An important aspect of our intelligent application is that each user can feed with reliable information, the clearest and most detailed description of a traffic accident and those involved as you can see in figure 4.

The test lasted a total of 4 weeks. Within this period, the users were able to register a total of 111 reports. There are two ways to get an overview of accidents in our intelligent application, in figure 5 you can correlate accidents by type and in figure 6 you can particularize an accident and if there were injuries or deaths coupled with geospatial data of it.

Based on the worldwide known Google maps, importing the data from the database, the incidents were added to a map to understand the location of the reported incidents according to the registered coordinates, as can be seen in the next figure 7, which describe each accident within the range of occurrence and its radius of affectation considering the time and degree of actors involved.

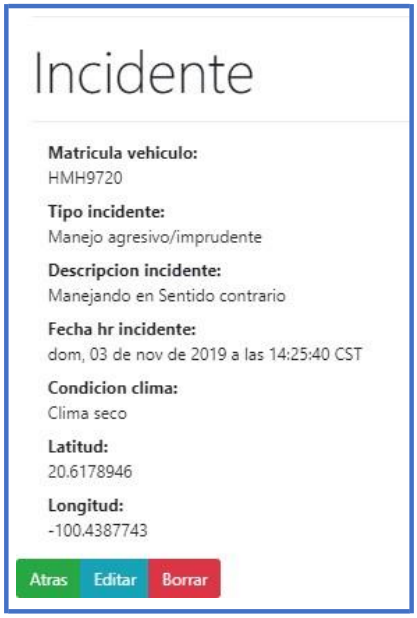


Fig. 6. Correlation of each traffic accident with relative information associated with vehicles involved.

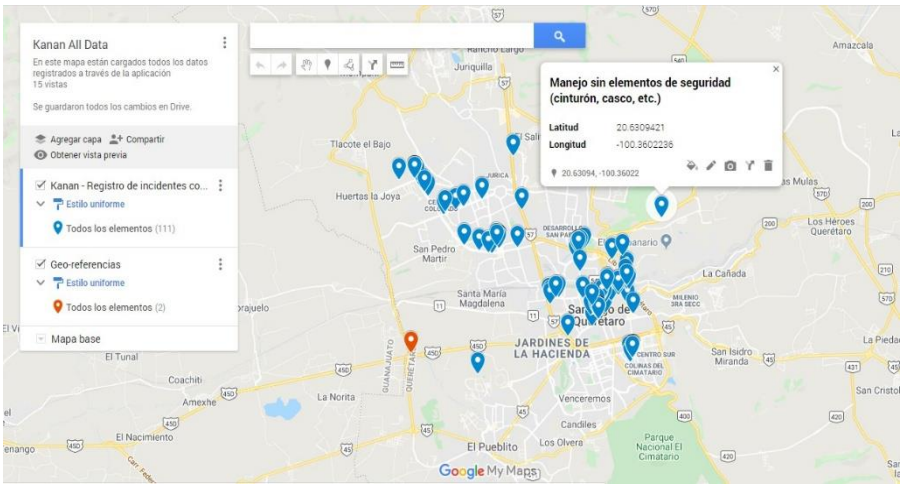


Fig. 7. Detail description of each traffic accident in a Longitudinal study to our intelligent application.

In order to do this, it is necessary to save the information in a spreadsheet with a CSV format, which can be used to automatically transfer the information to the Google platform. This is a simple way to understand the density of the participation of the user. In the map is very evident the areas where the users were reporting the incidents.

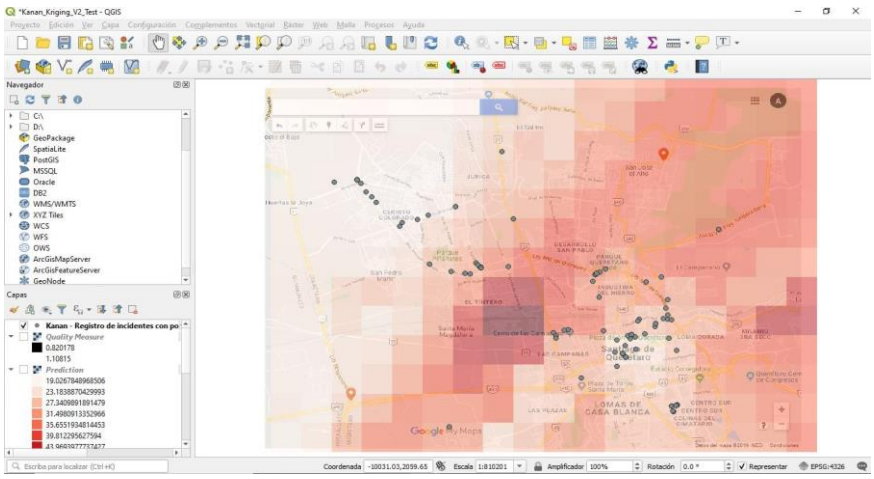


Fig. 8. A Kriging Model using correlative and transversal data to determine future traffic accidents.

But what exactly can be done with this information? We used a geo-statistical method named Kriging. This method uses the information from a geospatial delimited area, make an interpolation between registers with a similar origin or classification, to point out the areas that according to the analysis, are more susceptible (has higher probability) to an accident.

The analysis was done using QGIS, which is a free open-source platform for geospatial data analysis. In order to locate the data within a map, it is necessary to capture the map from any mapping source. In this case, we used the map from Google where we put the information at to visualize it. Then, selecting two specific points with its corresponding coordinates, there is a way to do the geo-reference. That way, once the reports are added, they will be located over a preliminary referenced map.

The reports, which are imported from the database, are loaded into the program which can be handled in different layers. Therefore, can be loaded different data to work within the same map. In this case, all the reports were loaded, and we ran the tool for Kriging analysis. The result can be seen in figure 8, we describe our proposed Kriging model to determine future numerical prediction.

According to the analysis, with such a few reports, the darkest area marked in the map is that with the highest probability of an accident. The way Kriging determines this information is doing interpolation of the different reports in different locations, and find in different directions the level of influence on of them can influence in some others.

4. Conclusions and Future Research

It is through all this validation process that we learned the potentiality of this system. The way that the tool could gather the information with just a few users and also

provides such a useful information with just a few reports, provides a very good taste of how good this proposal would work in a larger scale. The output of this exercise could be used, in this case, to focus the strategic measures to prevent accident in the conflict zones. If the measures are implemented in the areas where is predicted a higher probability of an accident, the impact of the measure will be higher. This provides a very good feeling about how useful and practical the proposed methodology can be.

With the use of this innovative application combining Data Mining and a Kriging Model based on a mobile dispositive is possible to determine the places where a traffic accident is possible to occur in Ciudad Juárez; by an alert sent to a mobile device with GPS, providing statistical information through a Web server that returns the level of rush hour in the area consulted [4]. The most important contribution is the possible prevention of future deaths in the city caused by traffic accidents. The future research will be to improve the visual representation of traffic problems with real on-time information through an Intelligent Diorama to specify an umbral of danger to an intersection of a Smart City. This design will bring common information to family members or social network close members.

Another possible extension will be to update our database using recent data from the local Transit Department's central server of security. Taking into account the 47,500 deaths related to traffic accidents in the last six years in Mexico, we think this innovative technology has promising application in another metropolitan cities in Latin America with similar problems of traffic, such as: Araguaiana, Blumenau, Joinville, Londrina, Manaus, Rondonópolis, Uberlândia. There exists a plausible application for motorcyclist since they face a higher risk of traffic accidents in Queretaro: 55 deaths out of 67 accidents in 2019. In addition, this application will be used as Recommender System when traveling to another countries [7] and explain different scenarios according time and location. In a future research we detail a prediction model using Random Forest to predict where will be occurs an attack of carjacking with basis on police department information, and using a hybrid model with another different Bioinspired Algorithm as Okapi Algorithm - This research group is researched about a novel paradigm related with Okapis, and proposes an idea about collective behavioral in a herd with skills different to each issue- or Wolf search algorithm with ephemeral memory.

References

1. Alonso, F.I.: Las coordenadas geográficas y la proyección UTM. Universidad de Valladolid, Ingeniería cartográfica, geodésica y fotogrametría (2001)
2. Andreu, R.A.: Estudio del desarrollo de aplicaciones RA para Android. Trabajo de fin de Carrera (2011)
3. Barrera J.O.: Sistema de Información Geográfica Móvil Basado en Comunicaciones Inalámbricas y Visualización de Mapas en Internet. M.C. tesis (2011)
4. Cáceres, A.: Sistemas de Información Geográfica. Instituto Formación Docente P.A.G. (2007)
5. Cruz, L., Ochoa, A., et al.: A Cultural Algorithm for the Urban Public Transportation. (HAIS´10), pp. 135–142 (2010)

6. Glass, S., Muthukkumarasamy, V., Portmann, M.: The Insecurity of Time-of-Arrival Distance-Ranging in IEEE 802.11 Wireless Networks. ICDS Workshops 2010, pp. 227–233 (2010)
7. Souffriau, W., Maervoet, J., Vansteenwegen, P., Vanden, B.G., Van Oudheusden, D.: A Mobile Tourist Decision Support System for Small Footprint Devices. IWANN, 1, pp. 1248–1255 (2009)

ISSN 1870-4069

Comparative Analysis of Disdain in Duets from Eurovision and Pop Idol Music: An Approach from Emotional Data Mining and Sentimental Analysis

Ana Elizabeth Salgado Arteaga[1], Daniela Beatriz Salgado Arteaga[2], Sandra Bustillos Durán[3], Carlos Alberto Ochoa-Zezzatti[3]

[1] Universidad Autónoma del Estado de Morelos, Facultad de Estudios Superiores de Cuautla, Mexico

[2] Salgado Arteaga & Asociados, Mexico

[3] Universidad Autónoma de Ciudad Juárez, México

alberto.ochoa@uacj.mx

Abstract. The present paper discusses a research related to analyze Sociolinguistics using Atlas.Ti a tool that mine information of the structure and content of speech related with a sample of 51 duets or bands which sung in Eurovision or conform icons from Pop Culture which develop a performance related with disdain reflex the feeling of a couple with problems with the purpose of discovering hypostasis and parataxis, which consists of relation in formal and informal use of language induced by the relation with another people interfering in their relationship, this phenomena has been documented recently, but with few detailed research with truly information, for this purpose we record speeches from different songs in a repository recreating these social behavior, to realize a detailed sociolinguistics analysis. We propose a model to represent emotional data mining to understand the relationships of each song with its relative music.

Keywords Sociolinguistics, social simulation, modeling of societies.

1 Introduction

Social Data Mining Systems allow the analysis of the society's behavior. These systems do that by mining and redistributing the information on computer files storing the social activity. Although, we generate two general questions to evaluate the performance of such systems: (1) is the extracted information of any value? And (2) is it possible to determine if a set of physical separated people can show a similar way of thinking about likes and preferences?

We made an analysis that provides positive answers for both questions. We live in an age plenty of information. The Internet offers endless possibilities. Web sites to experience, music to listen, chats rooming, and unimaginable products and services offering to the consumer an endless option varying in quality. People are experiencing

difficulties to manage the information: they can not and do not have time to evaluate the whole options by themselves, unless the situation seriously forces them to do that. In this paper we try to describe how two groups of couples sing about disdain from different perspectives about a sentimental relationship, these songs have similar discourses from a similar perspective.

A task to manage information which several internet users must do is "the subject management", searching, evaluating and organizing information resources for a specific subject sometimes Users search for professional interest subjects, some other times just for personnel interest. Our approach to this problem combines social data mining with information about sociolinguistics. In the daily life, when people desire forming part of a social group, without having the knowledge to chose among different alternatives, they trust frequently on the experience and opinions of others.

They look for advice in their ethnic-social group, perspective of a couple in love or not, familiar with certain likes and ways of thinking. For our research we need this information to understand how these sites on the web are populated and conformed. Social data mining can be applied to analyze the records generated on the web [5] (answering the question: Which are the most visited sites for the most of people?), online conversations [7] (Which are the sites where people purchase "thematic" things or for a community.

This paper is organized in five sections. In section one, we introduce our paper. In section two, we describe our Sociolinguistic approximation focusing in Social Data Mining. In section three we discuss the application of Atlas.Ti confirm the hypothesis of our research. In section four, we discuss the tests made to the analyzed information. In section number five, we discuss the results generated for the tests, and finally on the last section, we give the conclusions and future research of our research.

2 Sociolinguistic Approximation

Distinction between emotional grammatical and is not clear but it is possible to be conceited that emotional is pejorative and that thinks that the hypostatical style is superior. An analogy can be realized that "While a masculine oration usually is like a game of Chinese boxes, one fits within the other, a feminine one is like a Rep necklace them united by a thread of Greek is and other similar words", is for that reason that parataxis is common in British prose and the hypostasis are common in Renaissance prose, when a social group is affected by a historical success these tend to express more parataxis and less hypostasis.

2.1 Proposed Methodology

The motivation to make an approach by means of applications with Emotional Data Mining is based on previous research of Social Data Mining in this research area. This research area emphasizes the role of the collective analysis of conduct effort, rather that the individual one. A social tendency results from the decisions of many individuals, joined only in the location in where they choose to coexist, yet this, still it reflects a rough notion of what the researchers of the area find of what could be a correct and valid social tendency [6]. The social tendency reflects the history of the use of a

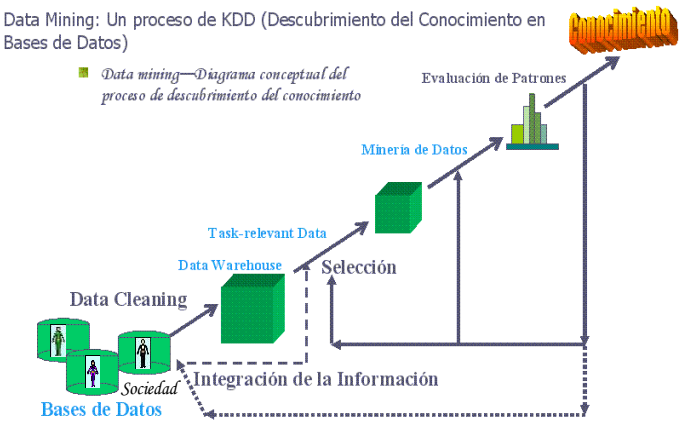


Fig. 1. Data Mining process. The society information inside a *Data bases* is cleaned and stored in a *Data Ware House*, then is mined by means of a loop back *selection* and *patterns evaluation* processes.

collective behavior, and serves like base to characterize the behavior of different ethnicity or socio-cultural cohesion [3].

3 System Development

The system will be able to analyze the behavior for the social networking of people involved in 51 songs performing by couples or bands in Eurovision in addition the same number related with Pop Culture Idols, by means of Atlas.Ti use, which has demonstrated being an efficient tool for searching hiding parameters that must be discovered [5]. The compiled information was analyzed to discover behavior patterns that share these individuals, and based on their gender, age, language or social cohesion and their interpretation about disdain, we determine if this behavior was an innate or induced tendency by their partners in each different song occurs in two samples from singers.

The name of Social Data Mining derives from the similarities between looking for valuable information in great data bases - for example: to find information of the tendencies of the society behavior in great amounts of stored Gigabytes – and mining a mountain to find a vein of valuable metals. Data mining automates the process to find predictable information in great data bases (See Figure 1). Questions that traditionally required an intensive manual analysis now can be directly and quickly answered from data [3].

4 Applied Tools

Atlas.Ti was used to analyze data. First, we proceed to develop a model that allows explain the behavior by the social networking, and how affects their speech style.

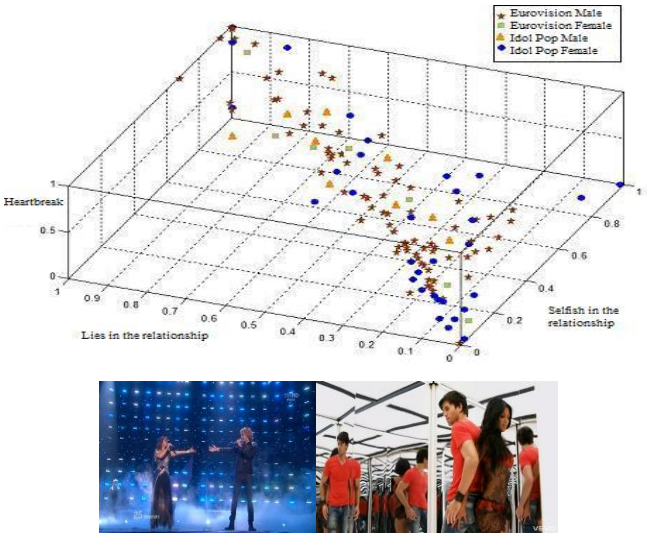


Fig. 2. Comparative analysis of Hypotaxis & Parataxis in a duet associated to two samples from Eurovision and Pop Idol singers.

Tables 1 and 2 discover the existent relation among hypostasis and parataxis parameters.

The male and female audiences held captive on the Eurovision music platforms and pop music idols are detailed in figure 2, and it can be seen that both platforms differ in terms of the female and male audiences they capture and in this respect it can be seen in Eurovision that the majority of the audience are men while the pop music idols are followed to a greater extent by women and the total of the captive audience on these musical platforms are encouraged by loving indifference in relationships by breaking their hearts, with the death of relationships and when dealing with selfish tendencies.

Although if we should put places Eurovision is followed mostly by men than women, and unlike the pop music idol is more attended by female audiences and to a lesser extent by men. This means that both men and women express their feelings in a very similar way and almost in the same magnitude, and yet men are the ones who express the most their feelings about break-up, death and love selfishness through duets or musical duets.

We found the indicator in all cases, especially when the people try to built a blog, a newspaper or follow a person in Twitter showed a higher hypostasis and banners, card gift or book in any language showed lowest parataxis regarding their participation in different situations of this social behavior, in figure 3, we show emotional data mining associated with a repository of Tweeters associated with these songs.

This can be explained by the use of informal speech of antagonism roll because they resist losing their love in a couple related with theirs feelings, and purchase decision is highly influenced induced by their community, in this moment many singers thinking that the people likely very much the songs with this thematic and the gender live a great époque, many specific linguistic tools and book and music projects related with disdain

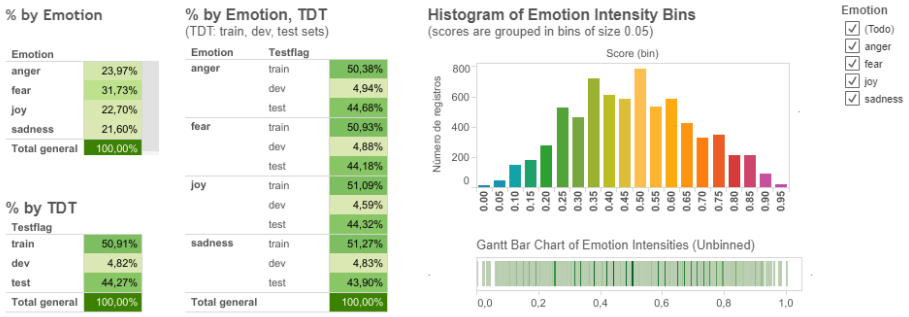


Fig. 3. Emotional Data Mining representation of these songs in Eurovision.

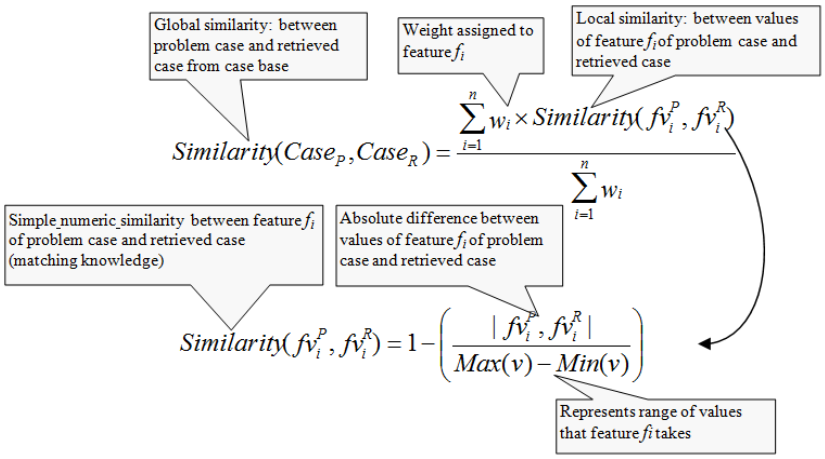


Fig. 4. Social distance between lyrics and emotional representation.

receive many acceptation in all world specially in Central Asia, East Europe and Latin America [8].

5 Results of Emotional Data Mining

We took in consideration the speech of a sample of: -51 couples participated in Eurovision and same quantity in Pop Idol from diverse countries- from two samples (Sample 1, Eurovision singers and Sample 2, Idol Pop singers) using their speeches in a social networking to identify different behaviors and explain the Imaginary collective of the people involved in this kind of songs (See Table 1). We applied a social similarity function to understand the closeness between lyrics and emotional representation of a song, as is possible see in figure 4.

The use of data mining in social aspects has demonstrated being key part to corroborate the linguistics tendencies of a group with common situations in different songs when the people suffer collectively, we found variations depending on the use of Sentimental analysis of emotions Speech, see Table 2.

Table 1. Distributions of demands by category and sort of the two analyzed samples. Sample 1 is associated with fans from Euorvision, and Sample 2 are fans club from Pop Idols duets or bands.

	Sample 1		**Sample 2**	
Category	F	M	F	M
N	13	27	28	39
Imperatives	12%	36%	26%	35%
Directives declaratives	5%	6%	7%	4%
Directives of Simulation	11%	4%	5%	6%
Interrogatives Directives	2%	0%	1%	1%
Interrogatives Postscripts	35%	16%	28%	19%
Joint Directive	15%	3%	11%	2%
Explosive Questions	2%	11%	4%	13%
Information Questions	16%	22%	17%	19%
Mechanisms of attraction of the attention	2%	2%	1%	1%
Total	100%	100%	100%	100%

Table 2. Contributions to the speech in a social network organized by kind of duet.

	Volume of Speech		
People	**Total of Emitted Words**	**Total of Turns**	**Average of words in turn**
Sample 1	788	127	4.9
Sample 2	992	108	7.2

In figure 5, is show the relation between duets, duration in radio and the size of bubble represents the quantity of original music from these duets in the time.

In this regard, it was found that Egypt (Africa-Asia), Argentina (Latin America), Sweden (Stockholm) and Belgium (Europe) are the countries that have the shortest radio time, with a 10 to 30 percent rotation, while the countries of Spain have the shortest rotation, Brazil (Latin America) and Kazakhstan (Central Asia) are among the countries with the highest average percentage of rotation in the radio (approximately 50 per cent), while Italy, France, South Korea, Malta, Montenegro (European countries) and Australia (Oceania nation), Colombia and Mexico (Latin America) have the highest percentage of rotation in the radio duets between 70 and 90 per cent.

Secondly, it was found that in terms of average time in months in the radio of duets, the countries with less temporality of two to four months are Peru (Latin America) and Georgia (Europe and Asia); while Egypt (Africa-Asia) and Japan (East Asia) are the countries with more months of duet music in the radio from 14 months onwards.

Thirdly, with regard to the representativeness of the countries in terms of the amount of original music by means of bubbles, it was found that Sweden (Stockholm) occupies first place, followed by Spain, the Netherlands, Belgium, France and Italy (European countries) and that the last places are occupied first by Mexico, followed by Colombia, Argentina and Brazil (Latin American countries).

An analysis of the findings shows that European countries have the highest production and rotation of music, while Asian countries have the highest rotation of duos on the radio and finally, emphasizing, Latin American countries have the highest rotation of duos on the radio and the lowest production of original music. Now,

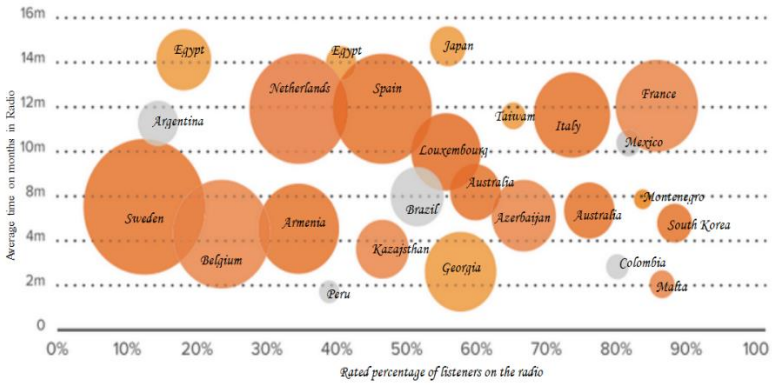


Fig. 5. Correlation between musical production, time on radio and quantity of original songs in a sample of societies of our study.

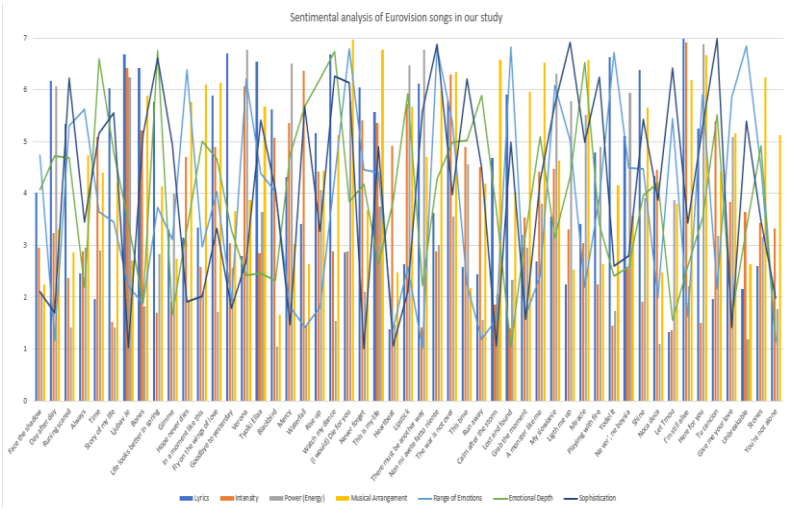


Fig. 6. Emotional representation of seven attributes in each song analyzed from sample of Eurovision (2000-2019).

emphasizing our country, Mexico can be detailed as having a rotation of listening to duets on the radio of approximately eighty-five percent (one of the highest worldwide), a temporality of durability of the songs on the radio of ten months (it is high in comparison with Japan that occupies the first place with more than fourteen months) and is the country with the least production of original music of duets; In this regard, it can be said that although Mexico does not produce high rates of music, it does have high demand for rotation and permanence in the radio, which means that people in our country are expressing their feelings through music and therefore the rotation and repetition of music is permanent.

In figure 6, we show the correlation between lyrics and different sentimental analysis according at age, social socioeconomical level, studies and life trajectories. In this

sense, the 51 songs of the duets that were chosen on the emotions of relationships were taken at Eurovision and linked to seven basic aspects: the lyrics, the intensity, the power/energy, the musical arrangement, the range of emotions, the emotional depth and the sophistication of the songs. With regard to the lyrics, it was found that these occupy first place to capture the attention of people (men and women) as they reflect the highest peaks, followed by emotional depth, range of emotions, musical arrangements, intensity, sophistication and ultimately energy as it is a matter of expressing couple feelings is not transcendental as in other music genres that may require combinations of high and/or low waves of power/energy that allow to transmit other types of emotions other than love, lack of love and couple problems.

6 Conclusions

With respect to the findings in the field, four important aspects can be concluded. The first is that the mining of emotional data is a useful instrument to evidence and analyze the feelings that people express through the adoption of certain styles of music.

Secondly, that there is a balance with respect to the transmission of feelings transmitted and/or expressed by men and women, however, there is a stronger tendency for men to express feelings of death, brokenness and selfishness in terms of love and/or lack of love evidenced by listening to musical duets.

Thirdly, European countries are the ones that have the highest musical production and rotation, while Asian countries have a higher rate of temporality of listening to music on the radio, while Latin American countries have a higher rate of musical rotation on the radio and less musical production; specifically in Mexico there is not a high musical production, but there is a high rotation and permanence of music on the radio, which means an important expression and externalization of feelings through virtual media. And fourthly, the attributes that people take most into account when listening to a song are: the lyrics, the emotional depth and the range of emotions, leaving out other attributes that they consider less important such as musical arrangements, intensity, sophistication and ultimately energy.

There are an important number of questions that deserve additional research. One will be to find new information sources to mine about the use of Cognitive Social Blockade Speech. With all these knowledges is possible predict a duet winner Eurovision for first time since 1995 may be is possible whim will be win is the duet from Azerbaijan which occurs in 2014.

In JESC'2020 two societies send duets Germany and Scotland in their respective debit, and they will be win. An area with great potential is the electronic usage of media, specifically, digital music [1].

In [6] is shown a system that learns of the user preferences based on the music listened, after songs are selected to be play on a shared physical environment, based on the preferences of the whole people present, this software has a narrative script to realize recommendations to another users in a free text [9].

Acknowledgements. We want to thank to ESC Radio by permit the use of Databases related with these people and their activities during the last year in the Web Radio.

References

1. Amento, B.: Specifying Preferences based on User History. In Proceedings of CHI'02 (2002)
2. Fiore, T.: Visualization Components for persistent Conversations. In Proceedings of CHI'01 (2001)
3. Demergis, D.: Predicting Eurovision Song Contest Results by Interpreting the Tweets of Eurovision Fans. SNAMS´19, pp. 521–528 (2019)
4. Pirolli, P.: Life, Death and Lawfulness on the Electrical Frontier in Proceedings of CHI'97 (1997)
5. Yazhou Zhang, Zhipeng Zhao, Panpan Wang, Xiang Li, Lu Rong, Dawei Song: ScenarioSA: A Dyadic Conversational Database for Interactive Sentiment Analysis. IEEE 8, pp. 90652–90664 (2020)
6. Toriello, A.: Beyond Recommender Systems: Helping People Help Each Other. HCI in the new Millennium, Addison Wesley (2001)
7. Winograd, T.: An Information-Exploration Interface Supporting the Contextual Evolution of a User's Interests. In Proceedings of CHI'97 (1997)
8. Cade, McCall., Lea, K., Hildebrandt, R.H., Blazej, M., Baczkowski, T.: Introducing the Wunderkammer as a tool for emotion research: Unconstrained gaze and movement patterns in three emotionally evocative virtual worlds. Comput. Hum. Behav., 59, pp 93–107 (2016)
9. Vijayaragavan, P., Ponnusamy, R., Aramudhan, M.: An optimal support vector machine based classification model for sentimental analysis of online product reviews. Future Gener. Comput. Syst., 111, pp. 234–240 (2020)

ISSN 1870-4069

Implementation of a Chromatic Recognition Model for a Scenographical Representation with Different Shades and Color Intensities for Color Blindness Actors

Christian A. Mejía Ramírez[1], Gustavo A. Toledo Nin[2], J. Fernando Estrada Otero[2], C. Alberto Ochoa Ortiz[2], Martín Montes Rivera[1], Julio C. Ponce Gallegos[3]

[1] Universidad Politécnica de Aguascalientes, Mexico

[2] Universidad Autónoma de Ciudad Juárez, Mexico

[3] Universidad Autónoma de Aguascalientes, Mexico

julk_cpg@hotmail.com, mc170003@alumnos.upa.edu.mx, martin.montes@upa.edu.mx, al158145@alumnos.uacj.mx, al157697@alumnos.uacj.mx, alberto.ochoa@uacj.mx

Abstract. At present some individuals are born with different capacities, some of these capacities are the perception of colors (color blindness); However, technology has advanced in medical matters, which recently through Artificial Intelligence (AI) techniques such capabilities can be improved and allow an adaptation of such individuals to participate in events where the perception of colors under certain lighting is essential, such is the case of the scenography in theaters and forums. In this research, a color labeling algorithm is applied in scenarios presented in theaters, said algorithm was proposed by Montes et al in 2019 [1], in this article methodologies of Genetic Algorithms (GA), Artificial Neural Networks are proposed (ANN), Particle Swarm Optimization (PSO) and Genetic Programming (GP); however, the GP methodology turns out to have better results compared to the others. That is why through this a color labeling is carried out in different scenarios where both actors with color blindness are presented, and in this way the lighting and color identification is not a limitation in the participation in said events.

Keywords: Color blindness, scenography, color labeling, color recognition.

1 Introduction

Throughout history, one of the senses that has had the greatest impact and relevance on human development has been vision, which is associated with the organs that enable it, called the eyes. This is responsible for seeing and knowing the size, shape, color, and location of everything around us, thus capturing images of our environment. To take care of our eyes we must read in good natural light, avoid being in front of the computer,

television, or electronic device for a long time. One of the main alterations in our eyesight is color blindness, a condition in which we cannot see colors normally. It is also known as color deficiency.

According to the National Eye Institute, 8% of men and 0.5% of women worldwide suffer from some degree of color blindness. One of the main characteristics of color blindness is when a person cannot distinguish between certain color [2]. For example, they often do not distinguish between green and red, and sometimes blue.

There are two types of cells in the retina that detect light. These cells are called rods and cones. The rods only detect light and dark and are very sensitive to low levels of light. The cones detect colors and are concentrated near the center of vision. There are three types of cones: some detect red, some detect green, and some detect blue. The brain uses the information sent by the cones to determine the color we perceive [2]

Color blindness can occur when one or more types of cones are absent, do not function, or detect a different color than normal. Severe color blindness occurs when all three types of cones are absent. Mild color blindness occurs when all three types of cones are present, but one type does not function well. It detects a different color than normal [2].

There are varying degrees of color blindness. Some people with mild color-vision deficiencies detect color normally in good light but have difficulty in dim light. Others cannot distinguish certain colors in any type of light. The most severe form of color blindness, in which everything is seen in different shades of gray, is rare. Color blindness usually affects both eyes equally and remains stable throughout life [2].

In some cases, color blindness appears as a congenital problem, but in other cases it can appear at some other stage of life. A change in the way the colors look can indicate the presence of a serious problem related to the condition [2].

2 Motivation

This work arises from the need that people with color blindness have to visualize colors in social recreation centers. In addition to helping them to have a better development in a work environment, this through the implementation of technologies that have been developed over the years such as color recognition models, also the effective lighting scale. It comes in turn with the sense of humanity in which we all have the same development opportunities, and in a romanticized way an environment given the characteristics of any human being, as one very similar and preferably the visualization of it.

3 State of the Art

Over the time, the problem about people who are found with some type of visual problem, even more specifically, in the detection and how to determinate colors, has led to the technology and implementation of systems applying the same system makes it greater today [3].

Firstly, making a hierarchical relevancy which is part of the application is way ends in the most relevant ass significative in terms of technology and in the opposite way ends in the most relevant according to the same applied in some type of analysis or research, it has been obtained that, after defining in a concrete way what color blindness is and how it influences the lives of those who present it, we decided to take place in the investigations and analysis to provide a better experience in the aforementioned individuals. So it is in the analysis of how the detection of the range of colors that is usually presented outdoors works, and later in the interiors of places, it has been suggested to the understand this trough variations in the color of objects in the images with daylight and surface reflectance models. In addition, the realization of the same it's been proposed through the relation of two approaches where both propose recognition of color, in which the first develops models based on the context of daylight illumination and hybrid surface reflectance, also predicts the color of objects based on the context of the scene. Other hand, the second method shows that the color of the objects can be non-parametric ¨learned¨ through classification methods such as neural networks and multivariable decision trees [4].

At the same time, the implementation of algorithms that have been integrated for the identification of patters developed to date are suitable for very specific cases, however, these end up somehow without covering all this existing field, since they are based on background substation to identify an output image. As an example, members of the ¨University of Engineering and Management ¨ have proposed as a method an algorithm that satisfies this, so in respect to the existing ones. In this, the main image (front)is implemented as a complement to the image subtracted from the background, in addition it creates an image with a relatively minimal loss of details [4].

In the same way, the implementation of course, it has been successful within the subject, has been applied in eastern countries, as it is in the case of japan, where the concern of inhabitants reside in an equitable environment where all you have to do is promote the infrastructure application culture as support is incredible [3].

In addition, laboratory studies to measure the visibility of lines under 100 and 1000 lx lighting conditions, information has been provided to reach an essential benchmark for the visually impaired when walking. The visibility of this it has been tested and its exact behavior that is based on the contrast of lines and backgrounds, in chromatic and achromatic color combinations. A total of 43 colors have been used to define target lines, and three achromatic colors for the backgrounds, therefore, a total of 129 color combinations were used. This is how the data from this experiment can aid in the design of walk lines on pavement surfaces, which can provide directions, warnings, and other helpful guidance to visually impaired. In recent years, improvements have been made to the living environments of the elderly and the disabled in countries around the world, along with developments that even become legal to facilitate independence in their daily and social lives [4].

In today's daily life we do not find directly related to technological equipment, the interaction with these equipment's is usually in the most cases of a single type called ¨interface¨ this interface does not allow interacting and managing the equipment. As well as performing the various tasks that it offers us, the interface being mostly visual represents a problem for people who suffer from an eye condition such as color

blindness, since there is this problem; investigations have been given by a solution or have helped improve the user experience [3].

So is the case of research carried out by the ¨National Chung Cheng University¨, which proposes a re coloration algorithm based on the eigenvector processing for a robust color separation under the transformation of the deficiency [3].

This re coloration is performed by means of proposed technique called ¨color deformation¨ (CW), this technique uses the orientation of the results of a simulation of color vision deficiency to deform the color distribution. In general terms, this technique performs a transformation of images in an RGB color space to a λ, Y-B, R-G based on the CIECAM02 model [5].

These algorithms can be completed with analysis of vision patterns, so as is known, humans tend to fix some specific points and regions in the images perceived during the first seconds of observation, these points and regions summarize the most important and significant parts of the observed scene. The ¨Istituto di Astrofisica Spaziale e Fisica Cosmica di Palermo¨ carried out studies on the differences in behavior of the human visual system, these studies are carried out by means of humans with normal vision and with color deficiency, in these tests the fixations will be traced. In the first 3 seconds of observation of color images for construction of fixation point maps, through these real fixation maps they performed an analysis of the differences between people with and without color vision [6].

This research resulted in a method to enhance the color regions of the image using a detailed color map of the segmented salient regions of the given image.

4 Genetic Programming (GP)

Genetic Algorithms (GA) are adaptive methods, generally these are used in parameter search and optimization problems, these are based on sexual reproduction and on the schemes proposed by Darwin on natural selection, specifically following the definition given by Goldberg. GA are search algorithms based on the mechanics of natural selection and natural genetics. They combine the survival of the fittest among sequence structures with structured albeit randomized, exchange of information, to constitute a search algorithm that has some of the genius of human searches [7].

In these types of algorithms, we start form an initial set of individuals, which are called population (this population is generated randomly). Each of these individuals symbolizes a possible solution to the problem posed, these individuals evolve based on Darwin's theory of natural selection, and adapt to a greater extent after the passage of each generation, this in order to get closer to solution required [7].

One of the variants or evolutions of genetic algorithms is Genetic Programing (GP), in which it preserves the same principle of genetic algorithms, genetic programming as genetic algorithms seek the solution to a problem, GP begins with thousands of programs randomly created computer models, which apply the Darwinian principle of natural selection, recombination (crossover), mutation, gene duplication, gene deletion and certain mechanisms of developmental biology. Therebefore, it generates an improved population over many generations [7].

The biological function of genetic programming is the same as that of genetic algorithms. For this reason, the operation is similar. The difference between one technique and another consists in the way of coding problems, which allows its use in series of environments where previously genetic algorithms could not be applied [7].

The origins of GP officially date back to 1992 after the appearance of the book entitled ¨ Genetic Programming¨ written by John Koza, where the term was coined and the formal foundations of this technique were laid, although there are previous works that without explicitly using the name of genetic programming can be considered as precursors of the matter [7].

5 Methodology

The representation of the color space for digital images, which come from a digital camera, is represented in the $\mathbb{R}^3$ space, where each color is represented by the triple (r, g, b) in RGB space. Sometimes AI techniques are used to separate a specific color [1].

An alternative for color classification is through the projection of the vector of a color in the subspace RG, GB and BR, in order to reduce the number of parameters for the optimization of six coefficients that satisfy the restrictions (1),(2) and (3), these are still related to the RGB native space [1]:

$$r \leq \alpha_1 g + \beta_1 , \tag{1}$$

$$g \leq \alpha_2 b + \beta_2 , \tag{2}$$

$$\gamma_1 \leq r \leq \gamma_2 . \tag{3}$$

In the article by Montes et al., of 2019, the inequalities generated by the GP for the color red are obtained, these restrictions are shown in (4, 5, 6), other colors generate a similar structure with this methodology [1]:

$$\frac{16}{sin(g)} - 6 \leq sin^2(g) + \frac{6}{b\, sin(b)} , \tag{4}$$

$$\frac{\cos(g)}{gr} - \frac{3}{\sin(g)} \leq \sin(g) - \sin(b) - \cos(g) + r^2 - 6 , \tag{5}$$

$$b - 9g \sin(g) - 5 \leq 7 \sin(r) - \frac{9g}{2b} . \tag{6}$$

The generated equations obtain similar efficiency results compared with human proposed equations, but the variations of luminosity can be supported by the algorithm, with the generation of new equations adapted to support light variation, and all these generated structures are highly parallelizable because r, g, and b are arrays with all the pixels of the image, obtaining one thread per pixel [1].

The labeling of the images was carried out using the algorithm proposed by Montes et al of 2019, using the results obtained by the GP and training for different illumination

Table 1. Symbols of ColorADD™ in for the real-time labeling assistant here proposed [2].

Symbol						
Color	Red	Orange	Yellow	Green	Blue	Purple

Fig. 1. Original Image 1.

Fig. 2. Image 1 as seen by color blindness.

Fig. 3. Image 1 labeled by symbols ColorAdd™.

Fig. 4. Image 1 labeled by color name.

Fig. 5. Original Image 2.

Fig. 6. Image 2 as seen by color blindness.

of the scenarios, with a population size of 400, a tournament size of 20 , a mutation probability of 0.275 and 1000 generations [1].

Some special symbols used to label the colors are shown in table 1, these codes were proposed by ColorAdd™ and can identify different colors and their luminosity, but despite these possible solutions, there is no definitive recognition [2].

Fig. 7. Image 2 labeled by symbols ColorAdd™.

Fig. 8. Image 2 labeled by color name.

6 Results

The processed images have a resolution of 960×720 and are in jpg format, some of these images are shown in Figures 1 and 5. In Figures 2 and 6 show the images as seen by a color blind. In Figures 3 and 7 it is shown in labeling by the symbols ColorAdd™ and in Figures 4 and 8 they are shown labeled by means of the name of the color.

7 Conclusion

In recent years, advances in AI allow the development of different products to identify and classify colors in people with color blindness. Although there is currently no cure, there has been a great advance in technology through AI techniques, so that people with such condition adapt and participate in cultural and recreational activities, such as works presented on stage sets.

In this research, the GP was used to carry out the labeling of colors for people with color blindness. These applications help users to identify colors in cultural activities where lighting and color are involved.

References

1. Rivera, M.M., Padilla, A., Gallegos, J.C.P., Canul-Reich, J., Zezzatti, A.O., de Luna, M.A.M.: Performance of Human Proposed Equations, Genetic Programming Equations, and Artificial Neural Networks in a Real-Time Color Labeling Assistant for the Colorblind. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11835, LNAI, pp. 557–575 (2019) doi: 10.1007/978-3-030-33749-0_45.
2. Neiva, M.: Color Add - Color identification system (2010) https://repositorio-aberto.up.pt/bitstream/10216/89233/2/168433.pdf#page=41.
3. Lin, H.Y., Chen, L.Q., Wang, M.L.: Improving discrimination in color vision deficiency by image re-coloring. Sensors (Switzerland), 19(10) (2019) doi: 10.3390/s19102250.

4. Iwata, M., Kitamoto, H.: A study on the visibility of chromatic and achromatic color lines to visually challenged people. Japan Archit. Rev., 3(1), pp. 99–112 (2020) doi: 10.1002/2475-8876.12128.
5. Ogawa, T., Oshiro, N., Kinjo, H.: Generating function of color information detection using genetic programming. In Proceedings of the 14th International Symposium on Artificial Life and Robotics, AROB, pp. 614–617 (2009) doi: 10.1007/s10015-009-0704-z.
6. Bruno, A., Gugliuzza, F., Ardizzone, E., Giunta, C., Pirrone, R.: Image Content Enhancement through Salient Regions Segmentation for People With Color Vision Deficiencies. 10(3), pp. 1–21 (2019) doi: 10.1177/20416 69519841073.
7. Eiben, A.E., Smith, J.E.: Introduction to evolutionary computing. In: Natural Computing Series, 28, 2pp. 17–20 (2015)

ISSN 1870-4069

Preventing the Trampling of Wildlife through a Heuristic Mechanism for Drone Swarm Auto-Organization in the Sierra de Huautla Reserve

Alberto Ochoa-Zezzatti[1], Gilberto Ochoa[2], Lina Aguilar[3], Alejandro Acosta[4], José Mejía[1], José Diaz[1]

[1] Universidad Autónoma de Ciudad Juárez, Mexico

[2] ITESM Campus Guadalajara, Mexico

[3] Universidad Autónoma de Guadalajara, Mexico

[4] Universidad Cuauhtémoc, Mexico

alberto.ochoa@uacj.mx

Abstract. Preventing the trampling of wildlife is a multidimensional problem that includes, among other things, a geospatial analysis of areas where the habitats of certain species of mammals in particular are found, which must cross between areas that have been isolated. A review of the literature has identified that an adequate proposal has been to build safe corridors under bridges, although the height and space of these must consider the space for a large herd. Deforestation has several causes, but wildfires and illegal tree cut, which are man-made, are the mayor. Current technology used to detect wildfires are: meteorological stations and satellite image and satellites. This last is a great option, but, from space, the fire detection is until it is large enough to be seen from orbital altitude. A drone can be used to monitor a forest looking for fire signs before the satellite observes it, but a single drone to cover a large acre surface is not optimal. A drone swarm with auto-organization capacity, equipped with atmospheric sensors that detect fire hazard conditions or even a fire in an early stage, needs to be used to optimize the area coverage. Implement a heuristic algorithm for drone swarm auto-organization applicable for wildfire alert and detection. Forest fires are a big environmental problem due they are mainly detected until they have burned some square kilometers. When these are detected at the developed stage, the fire will be difficult to contain. Some wildfires affect agricultural along as residential areas causing significant economic loses.

Keywords: Trampling of wildfire, heuristic mechanism, drone swarm auto-organization.

1 Introduction

According to Bala et al [1], prevention of trampling of wildfire, deforestation and promotion of afforestation have often been cited as strategies to slow global warming. Deforestation releases CO2 to the atmosphere, which exerts a warming influence on Earth's climate. However, biophysical effects of deforestation, which include changes in land surface albedo, evapotranspiration, and cloud cover also affect climate.

This research is presented in 4 sections. In the first section, an overview of the geographical area in Morelos where we are focusing this research, the trampling of wildlife prevention in the Mexican state of Morelos, which is very rich in natural resources. In section 2, we describe the natural resources, economic and social impacts of trampling of wildlife and forest wild fires in this region. Also, in this section we show what methods and technologies are used to detect wild trampling of wildlife and forest fires and how our research considers it. In section 3, we present the formalization of the problem and the proposed solution on how to use an autonomous, self-organized drone swarm to prevent forest wild fires. In section 4, we briefly discuss the opportunities this research represents for future applications regarding the use of self-organized drone swarm.

2 Forest Characterization in The Sierra de Huautla Reserve

Overview. In this section, the area in Morelos where this research is to be implemented, is briefly described, mainly ecosystem, as is seen in Figure 1. Morelos is located in Central region of Mexico, in the border with Puebla, State of Mexico and Guerrero.

It is conformed by 37 municipalities. The capital is Cuernavaca. On 2015 counting census, the population was 1,876,574. The state is very rich in cultures, which, co-habit with Tlahuica people in 4 native indigenous municipalities: Coatetelco, Hueyapán, Tetelcingo y Xoxocotla. A unique characteristic is that it is a 4,927 km^2 state which is composed by different ecosystems from Cold Mountain climate (similar to the ones in Sakha Republic) [2], to mountains, depressions and Forests [3]. It houses a very diverse flora and fauna as depicted in Table 1

2.1 Endemic Species in The Sierra de Huautla Reserve

The term species refers to a set of natural populations where the members that compose can reproduce each other, however, cannot reproduce with members of populations belonging to other species. The gene pool of the species maintains its integrity by biological barriers, for example, urbanization, predators, deforestation, ensuring reproductive isolation. The speciation is where a set of organisms, with a considerable number of members, are isolates geographically because the effect of people in its habitats or population ecologically of its original population, the variations that appear allow to become in a new species, this occurs when a specie has dismounted their habitat by different situations as trampling of wildlife or fire forest.

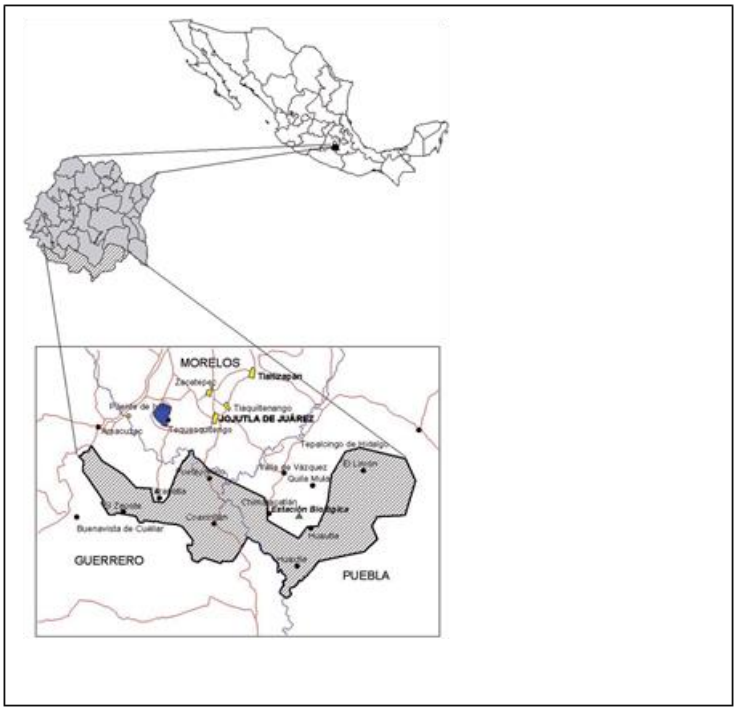


Fig. 1. The Sierra de Huautla Reserve in Mexico.

Table 1. Annual precipitation in Morelos's ecosystems.

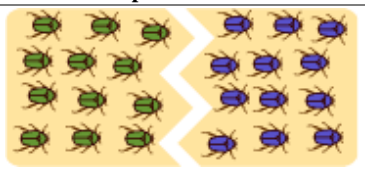
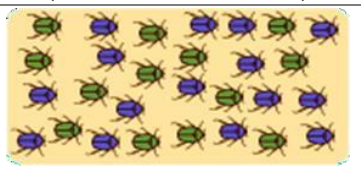
Climate Zone	Elevation (Mts)	Average Annual Precipitation (mm)	Months
Cfb	2,200	990	July - September
Cwa	1,300 - 2,200	700	July - September
BSk	1,200 - 1,500	475	August - October

During the speciation, reproductive isolation mechanisms (RAM) are introduced that prevent gene flow between populations belonging to different species, also occur the mechanisms of preggotic isolation that are those conditions of time or space that do not allow the formation of the zygote. If the mechanisms of preggotic isolation are not consolidated, interspecific mating could result, resulting in hybrid zygotes. In these cases, the postzygotic RAM, make impossible the development of the zygotes and that the hybrids achieve the adult state or they are infertile. In most cases, surviving hybrids are usually sterile.

The processes of speciation can be divided into two categories: speciation by divergence and instant speciation; Speciation by divergence refers to the fact that reproductive separation is formed gradually when a spatial or ecological barrier prevents gene flow between two sets of a population and instant or quantum speciation, the separation is established in a sudden form different kind of reptile.

There are four main models of speciation divergence shown in table 2.

Table 2. Types of speciation in an isolated population of issues in the Sierra de Huautla Reserve.

Type of speciation	Causes	Representation
Allopatric (alo = other, patric = homeland)	Geographically or ecologically separated populations.	 Allopatrinc Speciation. Source: (Caldwell & Collins, 2010)[12].
Peripatric	A population with few members separated at the end of another population with greater number of members.	 Peripatric speciation. Source: (Caldwell & Collins, 2010)
Parapatric	A population with a continuous distribution.	 Paripatric Speciation. Source: (Caldwell & Collins, 2010)
Sympatric	Within the range of the hereditary population	 Sympatric Speciation. Source: (Caldwell & Collins, 2010)

To perform the simulation with cellular automata, an endemic species that inhabits the state of Morelos, the squamous mesquite lizard was selected because it has been studied for several years by the speciation process that it presents. This species is protected by SEMARNAT, as they are threatened by the degradation of natural habitats due to urbanization, the division of spaces by the construction of motorways, intensive agriculture, grazing, burning, deforestation and tourist activities. In the state of Morelos, there is a severe impact due to the construction of Huexca Central power station. In the Sierra de Huautla Reserve, there is clandestine logging that causes depletion and contamination of water sources [4]. Poaching is pressing populations of important species.

The high incidence of forest fires reduces the opportunity for ecosystems to recover in time and form is therefore the importance of using drones that can help

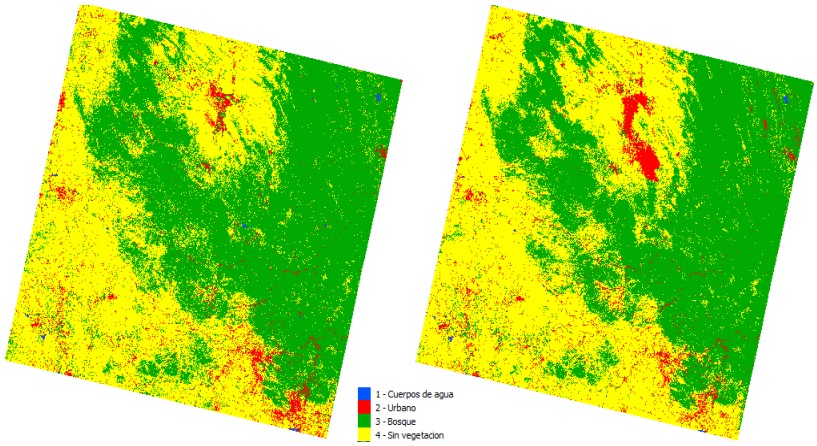


Fig. 2. Comparative map of the land use classification map of the years 2017 and 2020 in a section of Sierra de Huautla Reserve.

Fig. 3. Implementation of a proposal solution to adapt a drone to prevent trampling of wildlife and fire forest during a more widely time on Morelos' forests.

reduce this problem [5]. Considering the problem of wildlife run-over, it is necessary to establish the georeferenced accuracy of where the death situation occurs, determine the species and consider if the accident affected the ecology of any species population. In figure 2 is possible identify the reduction of habitats associated with specific species.

In figure 3, the use of a drone as part of a drone cluster can be analyzed to adequately determine the effect of wildlife runs, by specifying and validating whether a herd of a given species is attempting to migrate from one point to another in the space that makes up its habitat.

A decisive role in our research has been the implementation of an Industrial 3D printer in order to develop components of a drone cluster that allows to co-organize in order to detail searches according to the needs of the research to be carried out and to be able to diminish the effects of the fauna run over, especially in night and low visibility situations.

Table 4. Historical data of wild fires in Morelos, Mexico and its possible effect in trampling of wildlife.

Year	Quantity of fires	Burned Surface	Average Surface	Trampling of wildlife reported	Damage to the species' habitat	Reduction of population ecology
2010	1,057	18,505	17.51	2787	1.54	6.03
2011	625	10,560	16.9	2487	6.53	2.36
2012	1,153	17,216	14.93	1370	2.42	4.84
2013	842	10,704	12.71	1850	6.83	3.02
2014	697	29,316	42.06	781	6.03	2.58
2015	1,687	87,920	52.12	1362	3.79	5.31
2016	1,473	55,979	37.47	1035	3.44	5.02
2017	1,137	30,554	26.87	2069	1.02	4.86
2018	817	17,600	21.52	1124	2.88	4.56
2019	251	1,974	7.83	1980	6.78	2.60
2020	701	13,353	19.04	2997	3.97	6.21
Total Average	949	26,698	24	2493	4.29	4.77

3 Forest Wild Fires in Morelos State and Correlation with Prevent Trampling Of Wildlife

Overview. In this section, statistic data of Morelos forests wild fires that can give us an idea of the impact on the ecosystem as social and economics. Also, it's briefly described the technologies used to detect wildfires. Vegetation cover in forest ecosystems and as a consequence erosion and soil degradation. Historically, our State has seen affected by this type of casualties, the years of greatest occurrence were in the past from 1995 to 2004 and 2008, which showed a significant decrease to the year 2015, returning to present a rebound during the 2016, with a historical average of 870 annual fires between 1995 and 2020. The main causes of fires are still those related to Agricultural activities, slash and burn and the Crops, with a percentage that fluctuates between 25 and 60%. Regarding the type of ecosystem (CONAFOR 2015), the most affected is the cold temperate climate, followed by natural pasture. The smaller area report is the arid and semi-arid, as for forests there were no fires. The most common type of fire is the superficial fire. The year 2019 presented totally atypical conditions, since only Presented 252 forest fires, affecting a total area very large.

Surface recognition is a very common need like for agriculture, disaster relief operations, goods delivery, in another spatial issues. Depending on the area size, this could take a significant amount of time. For this, there's a need to use tools like a drone, but, also, this need so to have a mechanism to efficiently perform these activities and, mainly, surface recognition to optimize time. And resources depict need to coordinate the drone swarm to determine affectation in trampling of wildlife.

3.1 Case study Correlation between Fire Forest and Trampling of Wildlife

Wildfire detection and its effect on trampling of wildlife in early state case study for implementing a heuristic algorithm for drone swarm auto-organization poses some challenges:

A) Drone technology. Current drone technology is limited in capacity of taking long-range flights due battery life which is consumed in function of drone flight mainly. Adding a sensor data-logger and transmit data to a ground station represents a load to the battery life that has to be taken into account. Sensor and

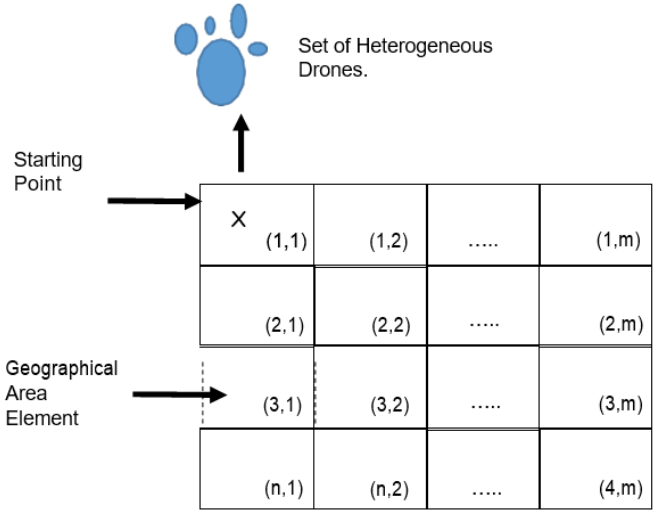


Fig. 4. Representation of the problem.

data transmission electronics does not represent a weight load in the aerial vehicle [6].

B) Radio signal. Data from the sensor data logging shall be transmitted to ground station. Due the forest geography, not all areas are covered by radio-signal. This can be solved by using some agents of the drone swarm serve as signal relay. This may represent a challenge to the auto-organization algorithm [7].

C) Altitude. Current drone technology is limited in the altitude it could reach depending on the orography where it will be implemented. This challenge is out of the scope of the auto-organization algorithm, but shall be taken into account due spherical coordinates will be used [8].

Since we're going to be working with very extensive surface of forests, these extension needs to be delimited by areas, which is the base on where the algorithm is going to define the priorities to visit by the drone swarm agents. The area to be monitored is considered in the algorithm. In the next section, it is described which are the already-known algorithms that will be referenced.

4 Formalization of the Problem

Overview. In this section, the problem we work on solving is defined and also, the drone swarm algorithm is described. We consider the problem of monitoring a large geographical area using a drone swarm to prevent forest fires which affect at many species associated with trampling of wildlife, as is possible determine in figure 4. The area to be monitored is divided into well identified sub-areas. A drone swarm is composed by a set of heterogeneous drones which are located at a starting point. Thus, the problem that we tackle is to create a schedule containing the assignment of

drones to geographical sub-areas for monitoring and detecting forest fires, such that the schedule completion time is minimized.

A scheduling system model for planning the visit of drones to geographical sub-areas consists of the following elements: the geographical area, drones swarm and an objective function for scheduling.

4.1 Large Geographical Area

The geographical area is denoted by *A*. Without a loss of generality, we assume that A has a square shaped area and does not contain any obstacle. The square shape was chosen for simplicity in the model. The area *A* can be divided into finite sub-areas forming a vector $A= \{a_0, a_1, a_2, .. a_{nm}\}$ of dimension n-by-m. For convenience, we consider a_0 as the base from which drones depart and return after complete their mission. We use *Geo*(a_i) to denote the geographical position of a_i on A.

4.2 Heterogeneous Drone Swarm

The Heterogeneous Drone Swarm (HDS) can be represented by a DAG *HDS: :(D, E).* *D* represents the set of heterogeneous drones that compose the swarm. *E* is the set of directed arcs connecting different pairs of drones, so $e (d_i, d_j)$ denotes a precedence that indicates that drone d_j cannot start its mission until d_i finishes its mission. For convenience, $Pred (d_i)$ denotes the subset of drones that directly precede d_i and $Succ (d_i)$ denotes the subset of drones that directly follow d_i. The entry drone is those with $|Pred (d_i)| = 0$ and the output drone are those with $|Succ (d_i)| = 0$. For simplicity, in these cases we consider the use of dummy tasks such that the dag contains only one entry and output drone. Remembering that the drones are heterogeneous, we represent the estimated flying time from the base at a_0 with EF*T:* :D × $A \rightarrow Int$, where EF*T* (d_i, a_j) denotes the time for a drone d_i to reach a geographical sub-area a_j. For simplicity, we consider that the flying time to return to the base at a_0 is the same than the time to reach a particular area from a_0. A drone can be assigned to different missions, but it can only perform one mission at time. Thus, at time *t* we consider $avail::D \rightarrow [0..1]$, which captures the availability of each drone at time *t*. Note that the time of the mission of a particular drone is given when it is working at full availability. $W(d_i)$ denotes the time for a drone d_i to execute certain work once it reaches a geographical sub-area. $Setup(d_i)$ denotes the setup time for a drone to start a new mission. We assume that information about the flying and setup time are provided in standard time units, compatible with our drone performance measures.

4.3 Scheduling Problem

Scheduling drones to geographical areas requires the consideration of four events: (a) The time at which the drone starts its mission. (b) The time for a drone to reach a particular geographical area. (c) The time for a drone to perform certain work once it reaches its geographical area and (d) The time for a drone to return to the base. Thus, we first need to predict the time at which a particular drone departs from a_0 to perform its mission to a particular sub-area and the time in which the drone returns to the base.

1. Set the drone flying time.
2. Set the drone setup time.
3. Set the drone work time.
4. Calculate DR_u for each dron by traversing the graph from the exit node to the entry node and keep the values in *L*.
5. Sort the drones in *L* in descending order of DR_u values.
6. Create a list *LSA* with the sub-areas composing A.
7. **while** there are unvisited areas in ***LSA*** do
8. Select the first sub-area a_m from *LSA*
9. **for** each available dron d_i (*avail*(d_i)=*1*) in L **do**
10. Compute *EDT* (d_i, a_m) value.
11. Compute *ERT* (d_i, a_0) value.

Fig. 5. The DERT algorithm.

We must first define two mutually referential quantities. *EDT*(d_i, a_m) is the *Estimated Departing Time* of drone d_i to a_m, it is calculated by:

$$EDT\ (d_i, a_m) = Setup(di) + max\ dj\ \in Pred(di)\{ERT(d_j, a_0)\}. \quad (1)$$

Setup(d_i) is preparation time for a drone to start a new mission. It is added to the result of the max block in Equation (3), which returns the maximum estimated returning time in which each drone in *Pred*(d_i) return to the base. This is calculated by *ERT*(d_i, a_0), which denotes the *Estimated Returning Time* of drone d_i to the base located at a_0 and it is calculated by:

$$ERT\ (d_j, a_0) = EDT(d_j, a_m) + \left(2 * EFT(d_j, a_m)\right) + W(d_j\ ). \quad (2)$$

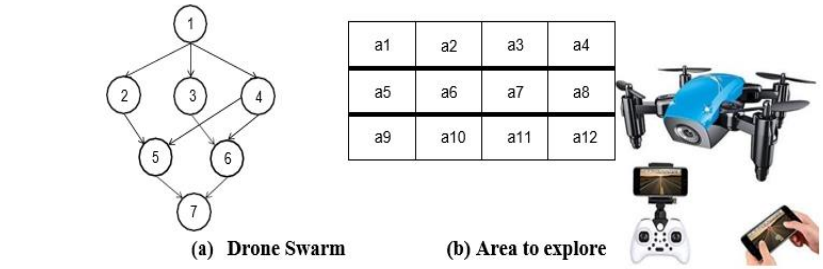
Once that all the drones have been scheduled, the estimated completion time of the schedule is determined by the estimated return time of the output drone. The estimated completion time is also known as the schedule makespan:

$$ERT(d_{output}, a_{m.\,n}). \quad (3)$$

The objective function for drone scheduling aims to create a schedule containing the assignment of drones to geographical sub-areas such that its makespan is minimized.

4.4 DERT Algorithm

The DERT algorithm is based on the well-known list scheduling approach. Our interest in this approach is to explore low computational complexity strategies and apply them to prevent and combat forest fires with the use of drones. Thus, the DERT Algorithm basically consist of two phases: The *drone prioritization phase* in which a priority rank assignment is set to each drone. The *geographical sub-area assignation phase* where each drone is assigned to that geographical sub-area which optimizes a predefined cost function. The DERT algorithm is shown in Figure 5.



Dron	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	12
1	7	3	9	4	9	4	3	3	1	2	9	9
2	4	7	18	3	2	7	7	2	7	8	7	5
3	8	2	19	4	7	1	2	8	4	4	4	1
4	9	9	17	3	2	7	9	1	2	2	3	5
5	4	8	10	5	9	9	8	8	9	1	4	9
6	5	5	9	8	2	3	5	4	5	8	8	4
7	2	3	11	2	3	3	3	7	3	2	7	3
8	7	5	14	4	7	5	5	5	5	3	3	4
9	3	4	20	3	4	4	4	4	2	4	1	4
10	1	5	2	7	2	5	5	5	10	7	3	2

Fig. 6. Proposed model for the location, monitoring, management and geolocation of wildlife runs to determine the effects on their herds and therefore on the ecology of their populations.

4.5 Drone Prioritization Phase

We use $DRu(d_i)$, an upward rank defined as the length of the critical path from drone d_i to the output drone. $DRu(d_i)$ is calculated recursively as:

$$DRu(d_i) = \text{avg}\,(FT_i) + max\ vj \in Succ(vi)\ (DRu(v_j)), \tag{4}$$

where avg(FT_i) is the average of the visit time for a drone d_i across all sub-areas.

$$avg\big(Ft(di)\big) = \sum_{k=0}^{nm} \frac{(d_i a_k)}{nm}.$$

5 Sub-area Assignation Phase

The DERT algorithm considers that a drone can be assigned to several missions, but it only can perform once at time. A mission involves to depart from the base a_0 to an assigned area a_m, perform a work once it reaches a_m and return to a_0. In our case, the work that a drone performs at a particular area is to monitor. The assignation phase

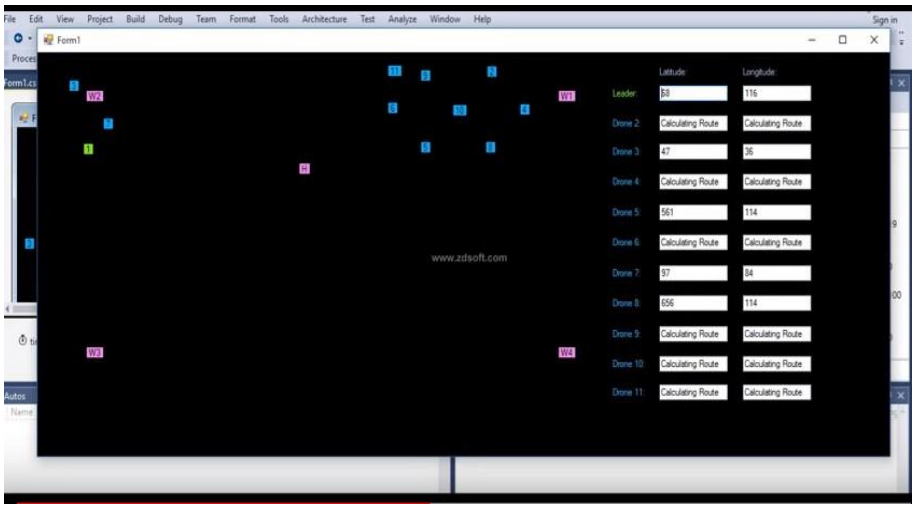


Fig. 7. Simulation software of our cluster of drones to prevent trampling of wildlife.

where a drone is assigned to a geographical sub-area offering the minimum estimated returning time, takes *O(d x e)* time complexity for *d* drones a *e* precedence's.

5.1 Example of Use in a Real Scenario

Using our drone cluster and the mobile application for its monitoring, it was possible to determine the effectiveness index associated with the evaluation of a given region, and to determine where wild fauna runs over, as can be seen in figure 6 and through the proposed model to determine the respective affectation to the ecology of populations of a given species.

By properly utilizing our mobile device to synchronize control of tasks associated with our drone cluster, we can properly identify the ideal wildlife management organization and achieve a predictive model that will reduce the number of wildlife strikes.

5.2 Simulation

This research still needs to be complemented with simulation and testing prior to implement it in a real environment. Beside the engineering challenges of a physical drone swarm, the developed algorithm described in section 3 will be simulated for fine tuning and validation. For this, the forest and wildfires will be modeled to replicate as close as possible the ecosystem were these drones will eventually be flying.

Using our intelligent application, we can visualize even in a night scenario if there is any wildlife run over and how these collateral damages associated with the decline of a particular species can be diminished.

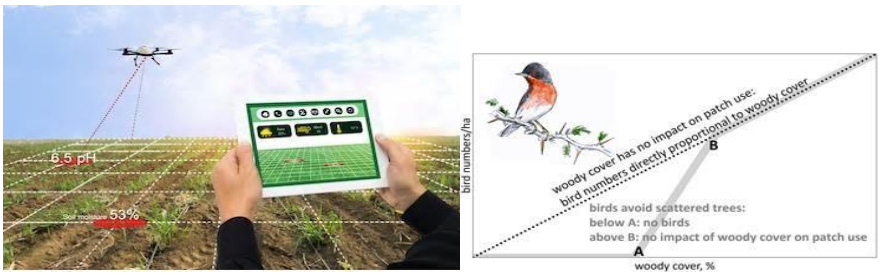


Fig. 8. A wood portion with illegal tree cutting identified by a machine learning model to determine the affection in a group of arboreal bird species.

6 Future Research

Overview. In this section, the future work for the drone swarm algorithm is described, the next challenges to be taken and opportunities, as Illegal tree-cutting detection in figure 8. It is important to adequately determine the amount of tree loss associated with illegal logging, in order to subsequently establish a level of actual impact associated with the future effects of specific changes and over time on the ecology of populations.

All around the world, illegal tree cutting is a big problem with several causes and big consequences to the global ecosystem as mentioned in the introduction of this research. This comes to account due one of the next works involving the presented heuristic algorithm for drone swarm auto-organization is to monitor forests to detect illegal tree cutting. Not only surveying forests with the flying unmanned aerial vehicles (UAVs), but equip this swarm agents with image recognition based on a machine learning model to be developed. The authors of this chapter are now starting to get involved on this endeavor.

References

1. Malhi, Y., Roberts, J.T., Betts, R.A., Killeen, T.J., Li, W., Nobre, C.A.: Climate change, deforestation, and the fate of the Amazon. Science, 319(5860), pp. 169–172 (2008)
2. ClimaTemps: Chihuahua Climate & Temperature (2017)
3. Köppen Climate Classification: Köppen climate classification (2017) http://hans chen.org/koppen/.
4. Jiang, P., Chen, Y., Liu, B.: Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks (2019)
5. Xiao, Y., Xing, C., Zhang, T., Zhao, Z.: An intrusion detection model based on feature reduction and convolutional neural networks. College of Information and Communication Engineering, Harbin Engineering University (2019)
6. Si, M., Tarnoczi, T. J., Wiens, B., Du, K.: Development of Predictive Emissions Monitoring System Using Open Source Machine Learning Library – Keras: A Case Study on a Cogeneration. Tetra Tech Canada Inc. (2019)

7. Hossain, L.: Deep Learning-Based Real-Time Multiple-Object Detection and Tracking from Aerial Imagery via a Flying Robot with GPU-Based Embedded Devices. Sensors, 19(15), 3371 (2019)
8. Si, M., Tarnoczi, T. J., Wiens, B., Du, K.: Development of Predictive Emissions Monitoring System Using Open Source Machine Learning Library – Keras: A Case Study on a Cogeneration. Tetra Tech Canada Inc. Calgary AB T2C 2X5 Canada (2019)

ISSN: 1870-4069
http://rcs.cic.ipn.mx

Centro de Investigación
en Computación