Obed Campos
Pablo Pancardo
José Adán Hernández-Nolasco

DYNAMIC FUZZY MODEL FOR DETECTING VERBAL VIOLENCE IN REAL TIME

Abstract *The crime rates in Mexico have been increasing in recent years; every day, there are reports on social media and in the news where assaults and verbal aggression by criminals can be seen. Public transportation units suffer from violence that authorities have not been able to reduce despite their efforts. This is why we have developed a fuzzy logic model that can adapt to almost any scenario thanks to the dynamism that we have implemented in each of its stages. We have obtained promising results that we believe will be of great help to the authorities for detecting the exact moment in which verbal aggression that is typical of a violent assault is happening in real time. This is a tool to help the authorities, not a substitution; we are simply making use of the latest technologies that are available to us. The goal of this paper is to provide a novel method for Mexican authorities in Mexico City in order to help the actual surveillance systems make faster decisions about whether violent assaults are happening or not.*

Keywords fuzzy logic, dynamic, verbal violence, real time

Citation Computer Science 23(4) 2022: 467–493

1. Introduction

In Mexico, 37.8% of the population has reported being a victim of some type of assault. According to data that was presented by INEGI (in English: National Institute of Statistics and Geography – the official institute that is in charge of keeping records on this and other matters), robberies or assaults on the street or in public transportation are the items with the highest incidences (according to the most recent report that was presented in 2019 [33], see Fig. 1).

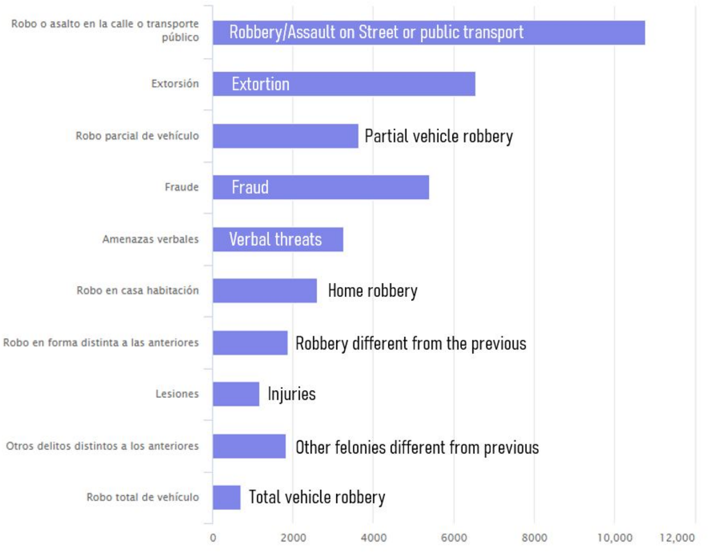

Figure 1. Criminal incidence INEGI 2019 graphic

This data is only a part of the huge criminal activity in the country, since only a few victims file formal complaints with the public prosecutor's office; the rest prefer not to do so for personal reasons or fear of reprisals from their attackers.

In 2018, the government of the state of Mexico installed security cameras (with microphones included) in those public transportation units that travel through the busiest avenues of the municipality. Thanks to this initiative by [22], hundreds of assaults on passengers have been recorded on video. This material is only available to the C5 (Center of Command, control, computing, communication, and quality of Mexico City) – a police security agency that is in charge of real-time monitoring of the activities of passengers on each of these public transportation routes.

Even though the technology has been implemented and is available for instant monitoring, it is very difficult to detect the exact moment when an incident occurs; there are many units that have such cameras and there are not enough security personnel to monitor the devices. As an effective security measure, panic buttons have been installed near the drivers so that they can press them in the event of an assault. At that moment, an alert is sent to the control center, and they notify the nearest patrol car to come and aid the victims inside the unit.

Due to the effectiveness of these buttons, criminals are aware of their existence and threaten the drivers in order to prevent them from alerting the authorities. This is a major problem, as the effectiveness has decreased over time, and some have stopped working due to physical or electronic failures.

It is worth mentioning that the audiovisual material is not available to the general public, as it is the property of the aforementioned C5. The material to which we have access has been shared through social media sites such as Facebook, Twitter, Instagram, and YouTube as well as the websites of popular newspapers in the country. For this reason, there is not a wide variety of these videos to properly test and generate an extensive database (or data set).

This represented an opportunity to generate a data set that contained key elements to assist in automated detection that is more effective than a panic button.

2. State of the art

There are works that have been done to help authorities and individuals detect crimes in the shortest possible time by using artificial intelligence algorithms and other techniques.

In the paper by Bautista-Duràn et al. [3], audio is analyzed by using two databases: one contains real audio files with violent content, and the other contains audio that is extracted from movie clips. Each of the samples has a duration of five seconds. In this paper, the method for analyzing data is by extracting audio features such as entropy and energy (among others) and use those features to compare whether the results match those features that are present n violent sound behavior.

To contribute to the databases that were built that contained violence data, [9] presented a recompilation of videos for violence detection at the 25$^{\text{th}}$ International Conference on Pattern Recognition (ICPR) in 2020.

Another research project that was carried out by Bugueño Sáez [5], consisted of detecting verbal aggression in audio files that were extracted from movie clips by using the Carletti model and the OpenSMILE tool (a tool that was developed by the authors specifically for this project) as well as the bag of words (BoW) and support vector machine (SVM) techniques. Bow & SVM are algorithms that are used to analyze word semantics; even though SVM is used in many other tasks, it has been proven to be good at these types of scenarios (achieving a detection rate of 86%).

Studies have been carried out to detect gunshots, screams, glass breaking, etc. [10, 11, 31, 37, 38]; however, these types of sounds are very easy to identify, as their pitches are very high and stand out among other sound levels.

In London, tests were carried out on public transport units in order to detect muggings. It is worth mentioning that these tests were done by installing surveillance cameras and analyzing video footage at police headquarters.

In Brazil, a team of researchers undertook the task of using a mobile application to record audio of an owner's phone in order to detect verbal violence [30]. This was done to help women, since it is very common for partners (usually men) to beat and abuse women in this country. According to study by Souza-Leal et al. [32], 607 sets of journalist notes were gathered from nine Brazilian news media outlets. This study stated that Brazil was one of the most dangerous countries for women, with the highest violent incident levels occurring during the years of 2013 and 2014.

The work by Santos et al. [29] presents a comparative study of deep learning that is applied for violence inside a car. The detection was based on an audio signal. The methodology that authors used for audio signal representation was Mel-spectrogram. They built a video data set and applied different deep-learning architectures to solve the classification problem.

After analyzing the existing state of the art, we did not find any work that has been done to detect verbal violence using fuzzy logic and applying dynamism to the model (as is the case of the proposal of this article). Therefore, we have great expectations about the functionality of the model.

It is worth mentioning that, even when visual data (video) is better for detecting violence that uses object recognition for guns, knifes, etc., we are only analyzing audio; this is due to the fact that audio analysis requires fewer computing resources than video. By analyzing words, we can detect violence even if a voice tone is calm.

3. Data acquisition & context variables

To carry out this research, data is required; this data needed to show violent events. We could have used public data sets that were made for this specific issue, but we decided to gather videos that were posted by social media users in order to take advantage of the violence wave that Mexico City was suffering. This led us to real assaults and not simulated ones.

The purpose of this recompilation is to classify acts into two types: violent, and non-violent. The detailed process of the data set will be covered in the upcoming sections.

3.1. Data acquisition

To obtain the data, a search was performed on the YouTube platform by using the following text strings:

- ASALTO COMBI ESTADO DE MEXICO (state of Mexico combi assault);
- ASALTO ESTADO DE MEXICO (state of Mexico assault);
- ASALTO CIUDAD DE MEXICO (Mexico City assault);
- ROBO TRANSPORTE PUBLICO CDMX (CDMX public transport robbery);
- ROBO VIOLENTO ESTADO DE MEXICO (state of Mexico violent robbery).

After obtaining the results, the next step consisted of observing the videos and searching for key elements such as the following:

- clarity of words and overall audio of video;
- most frequent words in robberies (examples: wallet, bag, bullet, iris, money, phone, and cellphone);
- violent or angry (altered) tone of voice.

Once a video met the requirements, its files were stored in a folder for later analysis.

The analysis consisted of identifying two extremely important variables (words and emotion); additionally, another variable determined the final result (location). The words and emotion variables were extremely important because they defined whether the contents of an audio were violent or not; if someone was blindfolded and heard violence, they would recognize it immediately because of the tone of the voice (emotion) and the spoken words. There are some dash-cam videos that show two assailants threatening a random taxi driver. They ordered him to start driving and insulted him in a tone that seemed like they were having a quiet chat [26].

For each file, the time stamp where the event started and the time stamp where it ended were identified. Subsequently, the total time was divided into five-second segments based on papers that yielded very good results (we based these on the time windows that were used by Siantikos et al. [31] and Bautista-Duràn et al. [3]). Each time segment was analyzed for the following parameters:

More significant variables

- The tone of the voice of the assailant (emotion). To categorize the level of emotion, we relied on the NOVACO Anger Scale (NAS) [Ref]; from there, we generated our own categories.
- The number of words mentioned in the segment that correspond to the previous analysis of the most frequent phrases or words.

Context variables (we will explain further)

- Time of the assault.
- Geographic location of the event.

With these parameters, the data set was built to train the fuzzy model that was developed in MATLAB using the toolkit that was included for fuzzy logic. This model is useful because of the toolkit that is specifically provided to design fuzzy logic models. We can have control graphically over the variables and rules, and they can be also modified via code. Although there are other choices (i.e., R or Python), we believe MATLAB is a good option; once the model is finished, it can be exported to compatible files that work as input for hardware like Arduino or another item from our list of selections.

3.2. Context variables

The context variables are external agents that are used to dynamically determine the amplitude of the values of the **words** and **emotion** variables (designed in the fuzzy model).

In every city, there are areas that are more dangerous than others (as is the case of the capital [22]). Based on this data, we determined that, when the geographic location is an area with high levels of incidences, the amplitude of **words** and **emotion** is higher. Consequently, the probabilities that the final result is a higher value are also higher. The same happens with the time context; the riskier the time, the higher the probabilities of obtaining high values in the final result.

There have been unofficial reports that show these external agents as factors to us; it is due to these that we decided to design these context variables. In Mexico, it is known that, when Christmas is approaching (during the months of October and November, and early December), criminals are more active than they are the rest of the year; this is due to the Christmas bonuses that working people receive during this period.

Other months that have shown greater activity than usual are March and May (Fig. 16). Even though we do not know in detail the reasons why these specific months show richer data, police reports are typically more numerous during these times.

4. Model development

The fuzzy logic toolbox that is included in MATLAB is very powerful and includes a user interface to make the configuration and construction of variables faster [34]. For each of the three variables that were defined (words, location, and emotion), we consulted relevant official data that helped us determine the design of each one. Likewise, we obtained reports from the Unit of Public Security to identify the location of the routes where the events occurred. The finished model took on the structure presented in the Figure 2.

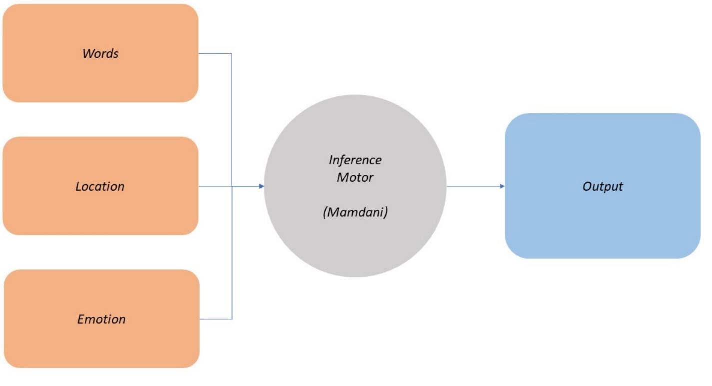

Figure 2. MATLAB fuzzy model

For the "words" variable, a trapezoidal structure was designed [36] that indicate the constancy of any words that are mentioned. Since there may be one or more words during a particular time interval, the maximum point is not unique.

The trapezoidal function is formally defined with the following equation:

$$f(x,a,b,c,d) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \frac{d-x}{d-c}, & c \le x \le d \\ 0, & x \ge d \end{cases} \tag{1}$$

This equation helped us design the variable. Other membership functions that were used include were Gaussian and triangular, which will be properly explained with their respective mathematical expressions.

With this, we have a membership function that is designed to help decide whether any spoken words contain a higher percentage of violence or not. The final design of the variable is shown in the Figure 3.

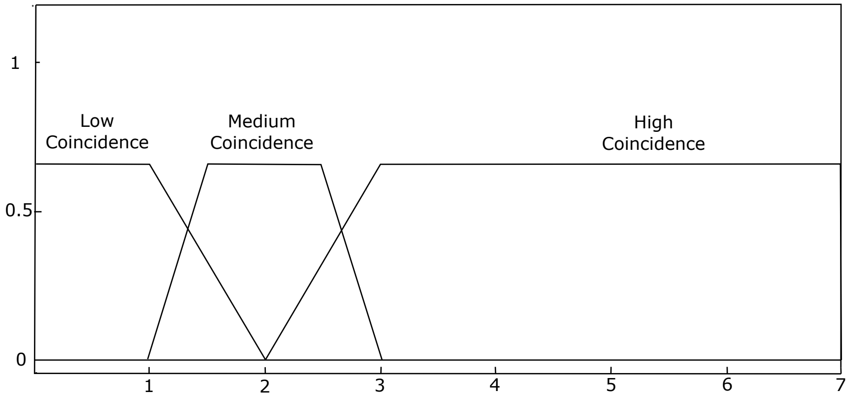

Figure 3. Design of words variable

Location is a variable that, even though it does not carry much weight in the decisions, is very important since it is part of the context. For example, if the words that are mentioned in a segment add up to three and the emotion has a value of "moderately angry" (we will explain this later in detail), the model output will be "violent", which is an average output value (not enough to raise an alarm).

On the other hand, if we add that the "location" variable to this same result and it has a value that approaches "very dangerous", then the centroid value in the output will move toward "very violent" (thus, raising alarms in police systems). Hence, the importance of this variable.

It is important to mention that we carefully analyzed different scenarios to cover all of the possible outputs and to not raise alarms when they are unwarranted. In this manner, we reduce the number of possible false positives in the equation. This variable was designed based on the Noticieros Televisa note [15], the most important TV News in Mexico. A transportation driver is interviewed, and he comments that all of his fellow drivers (including him) have identified the key points where assailants typically make stops in order to climb aboard and assault passengers.

For the design of this variable, we used a Gaussian model; this can be defined with the Equation (2):

$$G(x, m, \sigma) = e^{-(x-m)^2/2\sigma^2} \tag{2}$$

Therefore, the membership function design is presented in Figure 4.

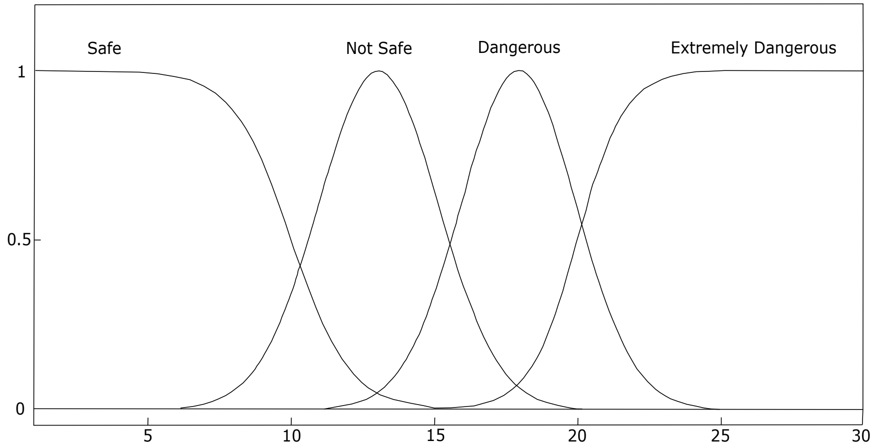

Figure 4. Design of location variable

One of the most important variables is undoubtedly "emotion", as it is a value that carries a lot of weight in the design of the rules for the model. If the number of words is considerably high (more than two) but the emotion does not denote that there is a startle in the voice, then we can understand that the situation that we are facing may not be an assault. Many people say such words when they are explaining a situation to friends or playing a joke, not necessarily justifying a criminal incident. On the contrary, if the emotion is "very angry" and the words pass up the threshold, then it is highly likely that the audio that is analyzed by the model is part of an assault.

To develop the design of this variable, we based our work on the study that was carried by Rodellar-Biarge et al. [27]. In this paper, biometric features were studied that can generate modifications on voice production (such as glottis movement or biomechanic parameters) that can produce stress in conversations where studied subjects defend ideas that go against their beliefs or feelings.

A triangular membership function can be defined as follows:

$$\Lambda(x,a,b,c) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ \frac{c-x}{c-b}, & b \leq x \leq c \\ 0, & x \geq c \end{cases} \tag{3}$$

The design of the emotion variable is presented in Figure 5.

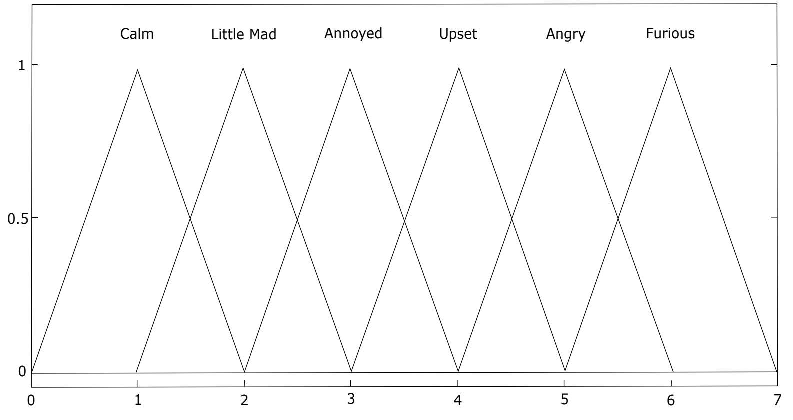

Figure 5. Design of emotion variable

The output of the model is the variable (see Fig. 6) that defines to which group an obtained value belongs (by calculating the centroid [25]).

The combination of values that is obtained in all of the variables by applying the **AND** connector of the Mamdani controller shows the output value [13] that we introduce as an input vector for the model.

This vector is formed by the values that are registered in the data set that we built by gathering the content of the videos and extracting the data and splitting each video into five-second windows (we call windows to a period of time).

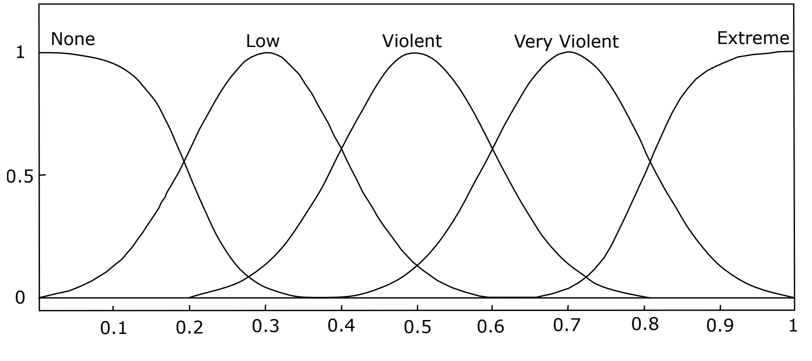

Figure 6. Design of output variable

A single video file can contain several windows (as can be seen in Figure 8). There is a column that is labeled **file**; this means that we are talking about the same video file each time that you see the same file value in the data set but that a different time period is inside it.

```
00:05   21-ago-18       [0;3;-1]        02
15      No_Enojado      -1      1       PUTA    9       00:40   00:45
00:05   21-ago-18       [1;9;-1]        02
15      No_Enojado      -1      0               9       00:45   00:50
00:05   21-ago-18       [0;9;-1]        02
1       Moderadamente_Enojado   2       7       CHINGARON, MADRE, PUTO,
TELEFONOS (2), CARTERAS (2)     22      00:32   00:37   00:05   17-oct-18
[7;22;2]        01
1       Enojado 3       5       TELEFONOS (2), CARTERAS (2), MATAR      26
00:37   00:42   00:05   17-oct-18       [5;26;3]        01
1       Enojado 3       5       TELEFONO (4), PLOMAZO   22      00:42
00:47   00:05   17-oct-18       [5;22;3]        01
1       Enojado 3       5       TELEFONO (5)    27      00:47   00:52
00:05   17-oct-18       [5;27;3]        01
1       Muy_Enojado     4       4       TELEFONO, VERGA, MOCHILA, PLOMAZO
19      00:52   00:57   00:05   17-oct-18       [4;19;4]        01
1       Poco_Enojado    1       0               19      00:57   01:02
00:05   17-oct-18       [0;19;1]        02
1       Enojado 3       4       TELEFONO (2), BOLSA, MOCHILA    22
01:02   01:07   00:05   17-oct-18       [4;22;3]        01
1       Enojado 3       3       MOCHILA (2), TELEFONO   27      01:07
01:12   00:05   17-oct-18       [3;27;3]        01
1       Enojado 3       1       MOCHILA 17      01:12   01:17   00:05   17
-oct-18 [1;17;3]        02
1       Muy_Enojado     4       2       PUTA, MADRE     18      01:17
01:22   00:05   17-oct-18       [2;18;4]        01
1       Moderadamente_Enojado   2       7       CHINGARON, MADRE, PUTO,
TELEFONOS (2), CARTERAS (2)     8       00:32   00:37   00:05   17-oct-18
[7;8;2] 01
1       Enojado 3       5       TELEFONOS (2), CARTERAS (2), MATAR      9
00:37   00:42   00:05   17-oct-18       [5;9;3] 01
1       Enojado 3       5       TELEFONO (4), PLOMAZO   6       00:42
00:47   00:05   17-oct-18       [5;6;3] 01
1       Enojado 3       5       TELEFONO (5)    9       00:47   00:52
00:05   17-oct-18       [5;9;3] 01
```

Figure 7. (Selected text) data to introduce as input in model

In Figure 7, we selected the vector that will be introduced to the model in MATLAB to obtain an output value. In Figure 8, we can see the vector that was introduced to the system that coincides with Rule 24 of the model (highlighting the group "little violent" in the output). The vector that was selected from the image was formed by taking the values of the variables and then built with a simple concatenation formula in order to save some work and apply some automation to certain tasks.

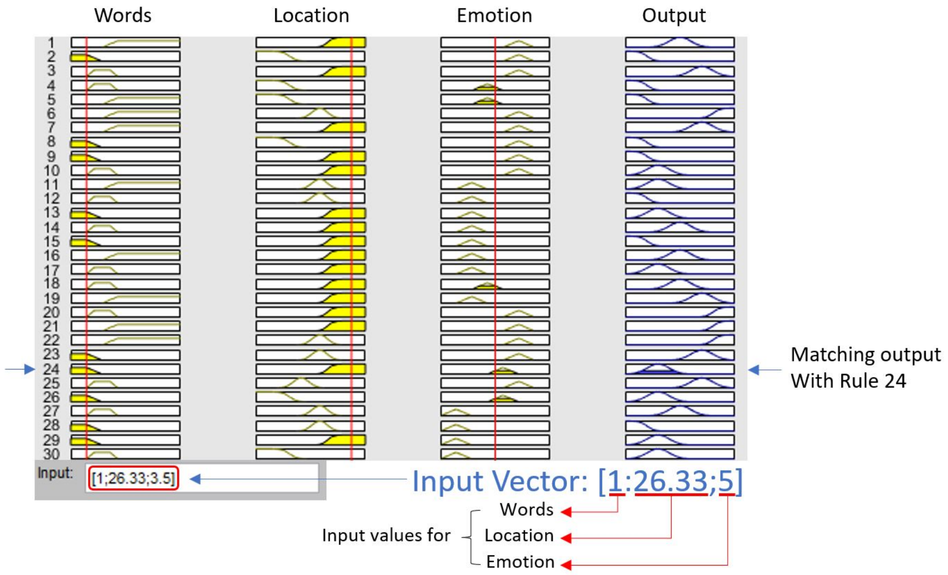

Figure 8. Input vector introduced to MATLAB model

In the image above, we can observe that the configuration of the rules and the input vector gives an output that uses the Mamdani method for fuzzy models. This method uses an AND connector between rules:

- example of simple rule:

> **IF:**
> Words = 4 **AND** Location = "Extremely Dangerous" **AND** Emotion = "Angry"
>
> **THEN:**
> Output = "Very Violent"

- another example:

> **IF:**
> Words = 1 **AND** Location = "Safe" **AND** Emotion = "Calm"
>
> **THEN:**
> Output = "None"

In Figure 9, we can see those results (output values) that show the combining rules of the Mamdani controller that are enclosed in the red rectangle (the resulting value is called **Centroid**).

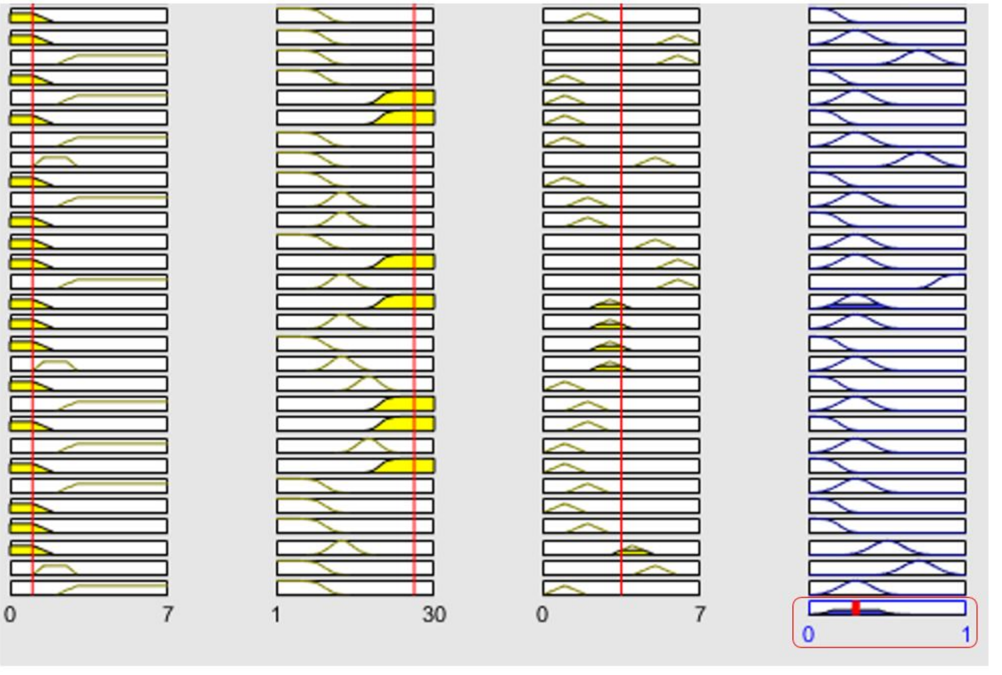

Figure 9. Centroid value

4.1. Making model dynamic

Once the input and output variables have been designed, the main contribution of this work is to add dynamic behavior to each of the stages of the model (i.e., dynamic inputs, an inference engine, and output).

Based on the data that was obtained from work by Naik et al. [17], we had an idea of how to make dynamic rules for the model. Naik et al. proposed a fuzzy rule interpolation (FRI) technique that consisted of modifying a ***snort.config*** file to interpolate some rules and make them dynamic. This is a good solution for systems that do not require real-time tasks.

We cannot use this proposal because of the fact that we need some configurations to be adjusted without having to stop the compiling of the system. Due to this fact, we had to implement our own solution in order to achieve the automation of the device.

In another work, we proposed a dynamic fuzzy system that changed the form of the variables (i.e., trapezoidal to triangular) depending on the requirements of the model:

> The presented approach consists of a fuzzy system where the membership functions can have dynamic transformations (according to the contextual variables that influence them) to have a model that adjusts in real time. The membership functions dynamism is achieved because the form in the sets can be transformed [19].

The first stage of the system is made up of the input variables; this is the part where the data that will be processed in order to obtain an output value are introduced. The system consists of three input variables: words, emotion, and location within the route.

Words variable

To determine the design of the variables, we compared multiple works [4, 6, 7, 21, 23, 35] and analyzed 38 videos in order to extract those words that are most frequently used by assailants the time of an assault. After counting, we found that there are about seven words that are frequently used and present in most cases.

Due to this, the words variable has a range of [1–7]; the groups that belonged to this variable were taken based on the number of words that were mentioned in each five-second segment for each video. That is to say, each video was divided into five-second fragments; therefore, if an assault has a duration of 3 minutes, 36 segments or time windows can be extracted.

Location variable

Each public transportation route travels a certain distance in kilometers. During each route, each unit passes through different areas (with some being more dangerous than others).

According to reports that were presented by the public safety office [18], a map was drawn up that showed the most dangerous areas through which each route can pass. This data was collected by running surveys and interviews among passengers and unit drivers.

In Mexico, criminals are intelligent enough to develop strategies in order to stay ahead the movements of the police. They have a community as well as contacts inside the authorities that let them know where patrols are located and where not to perform a robbery.

Also, they have WhatsApp groups where their friends publish information on when passengers could carry more money than normal due pay days or when a passenger just leaves a bank with loads of money (yes, they have contacts inside banks as well). All of these scenarios are a small part of what is called organized crime in Mexico; it is the smallest link.

The more insecure areas of Mexico City are detailed in the image below (Fig. 10).

With this in mind, we decided to assign levels to the variable that forms the groups of belonging (safe, not very safe, dangerous, and very dangerous) and a range of [0–30] given that, on average, the route of each unit is around 30 km in length. It is worth mentioning that the values in terms of ranges can be modified depending on the needs of each scenario or situation [8, 14, 24].

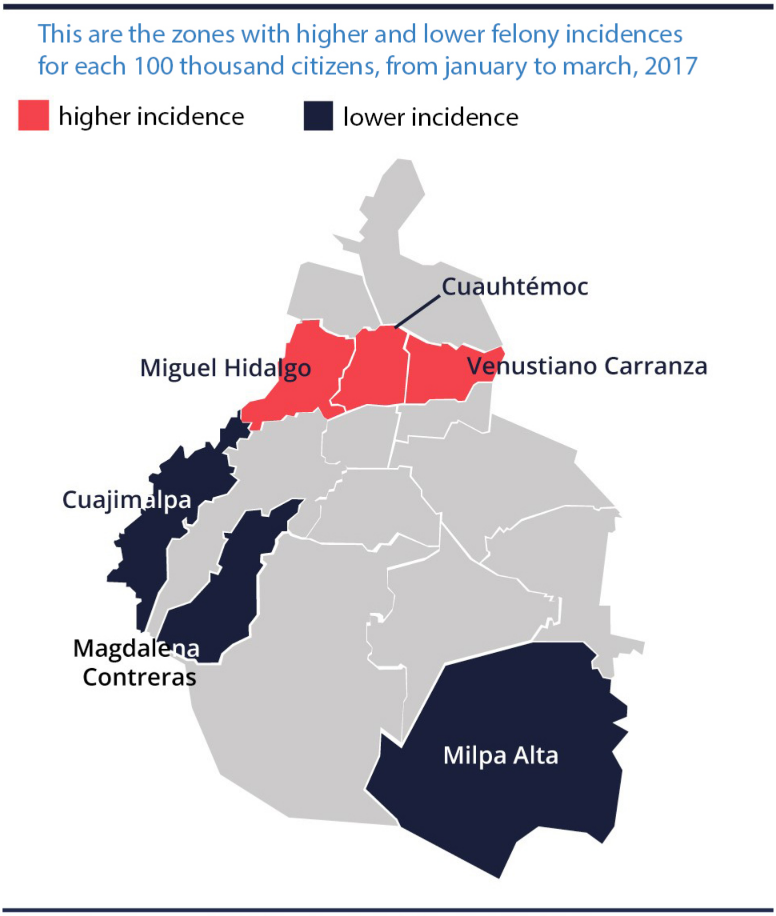

Figure 10. Map of dangerous zones in Mexico City [18]

Emotion variable

The design of this variable is based on the NOVACO Anger Scale (NAS) [2], which is a scale that arose from a psychological study that was conducted in 1998 on criminals in a prison in which levels of anger were determined and categorized on a scale in [16].

We decided to add one more level to the scale since the analysis must take time windows in which there is no presence of any emotion into account.

To make the inputs dynamic, we designed variables that will change their amplitudes depending on the context variables. Recall that these variables are the month of the year and the geographic location.

To achieve this, we decided to assign a "status" to the combination of context variables (that is, as a guide table to determine the change in the amplitude of a variable, see Tab. 1).

Table 1
Guide for amplitude control on each variable

Month.	Location	Status	Amplitude
Range 1.	Low Danger	Low	0.3
Range 2.	Low Danger	Low	0.3
Range 3.	Low Danger	Medium	0.6
Range 1.	Medium Danger	Medium	0.6
Range 2.	Medium Danger	Medium	0.6
Range 3.	Medium Danger	Medium	0.6
Range 1.	High Danger	Medium	0.6
Range 2.	High Danger	High	1

The amplitude refers to the maximum value that the variable takes on the x-axis (see Fig. 11). For example, if the amplitude takes on a value of 1.3, this means that this is the minimum value to which the vector should be set.

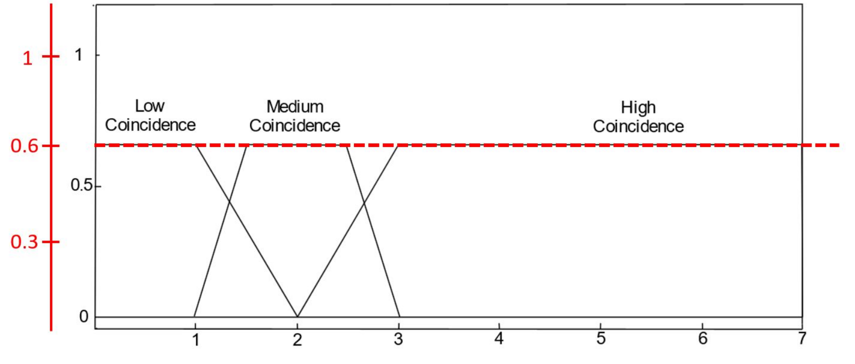

Figure 11. Sample of amplitude variation

By simply modifying the amplitude value, the system is automatically transformed into a dynamic model, as the system will change its input values depending on the context conditions and the result will different for each variation.

To achieve this result, we modified the MATLAB algorithm that handles the behavior of the input variables. By means of the input vector, we made a modification using a custom membership function for each of the shapes of the functions (trapezoidal, triangular, sigmoid, and Gaussian). What we did was to multiply the function by the last parameter that was entered into the vector. As an example, we have that an input vector for a trapezoidal function consists of four parameters (as is shown in Figure 12).

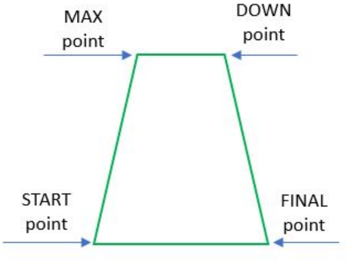

Figure 12. Points for trapezoid function composition

From the resulting vector from one of the functions, we can observe the following: [1 1.5 2.5 3] (whose description can be found in the Figure 13).

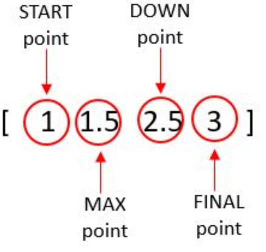

Figure 13. Trapezoid vector configuration

To make a more complete contribution, however, we went further and decided to apply this same dynamism in the inference core (that is, in the rules that control the result in the output variable). In order to control the dynamic form that is required in the rules, we cannot apply the same scheme of the input variables since there is no amplitude value in the rules.

Inference core (rules)

Rules are an important part of the model's inference core, as they determine the output value of the model. The non-dynamic rules can be very predictable since their very operation denotes in which group an output value will be. However, we can play with this value and add more or fewer values that belong to a certain group by dynamically modifying the weight of that rule (that is, each rule has a weight value that is set to 1 by default). This value is modifiable, and it ranges from 0.1 to 1 (which means that we have nine possible modifications for a single rule).

In our model, we decided to rely on the context variables to modify this value in the weight of each rule. This means that, depending on the resulting value of the context, we can modify the weights of certain rules in order to obtain a variable output value and, therefore, a dynamic behavior in the inference engine by means of its rules.

Output

The output of the model is the value that will indicate the probability of belonging to an output group. In this case, the value will indicate how likely it will be that a certain audio contains violence. For example, suppose the output value is 0.78; this indicates that it is a 78% probability that an assault with violence is occurring.

In this variable, the dynamism happens by modifying the number of groups according to the geographic location. In Figure 14, we can observe that there are three groups; this means that the location is less dangerous as compared to a variable with the groups of five (that is, some areas are more dangerous than others in a city). Therefore, the output variable will contain five groups instead of three when a user is in a more dangerous area (cf. Fig. 15).

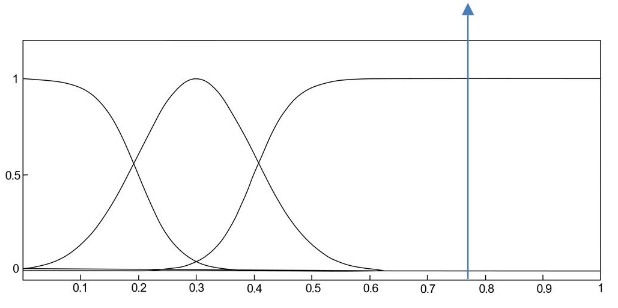

Figure 14. Graph for output value

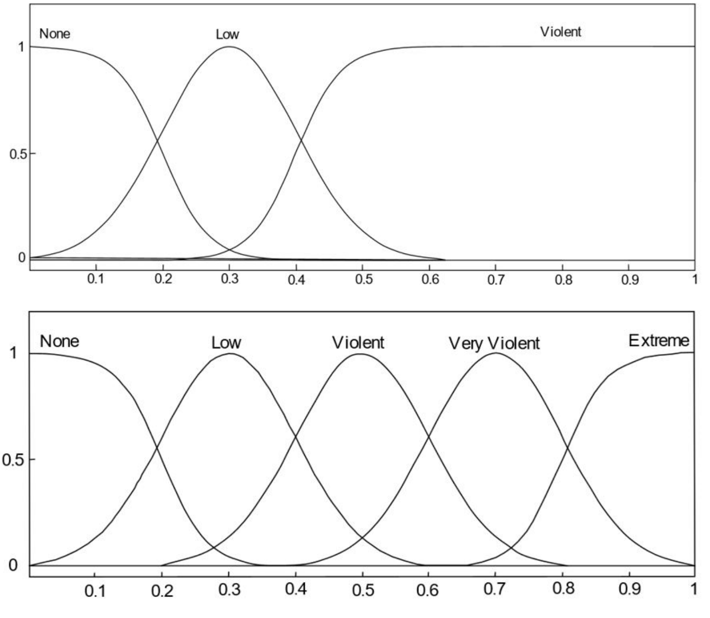

Figure 15. Output variable (groups of three and five)

Context variables

Context variables are external elements that determine the dynamic behavior of the input variables, rules, and output variables [1, 12, 20]. As an example, we have the example mentioned above. The number of output groups depends directly on the area or geographical location; in this way, we obtain a dynamism in the output variable.

As is the case of the output, there are factors that determine the dynamic behavior of the input variables and rules. These factors are represented in Table 1; they dictate the dynamic behavior of each of the phases of the fuzzy model.

These context variables were designed by us based on reports and statistics that were published by official security agencies of the city and state of Mexico. Likewise, the public security secretariat represented crime incidences by month and year by means of a graph; from this, we obtained the data that represents the MONTH context variable. The graph is as follows (cf. Fig. 16).

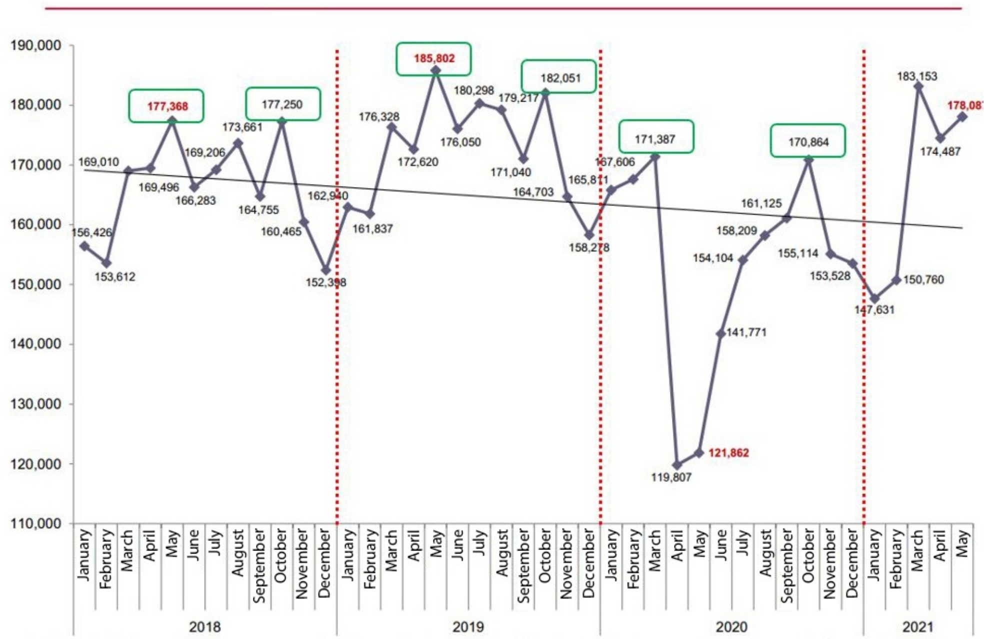

Figure 16. Accumulated criminal incidence reported by Public Safety Office

Let us note that, during the months of March, May, and October of the previous years, there was a peak in the graph; this indicates that these were the months in which there was the more activity on the parts of assailants. This study is of vital importance, since it is real data that was obtained from citizen complaints that were filed at the police and public prosecutor's offices; this is data that does not need validation, since the same governmental body officially validates it.

5. Experiments & results

Training

For the experimentation stage, the resulting data set was divided into two parts: 70 percent for the model training, and the remaining 30 percent for the testing stage. The data set had a total of 548 records, of which 383 records were randomly taken for the training stage and the remaining 165 records for the testing stage.

In the training phase, each of the resulting vectors was input to the model that was developed in MATLAB. If the output value matched the correct class in the data set, the record was marked as "trained"; if the resulting value was not categorized within any group in the output, then the rule was created and the model was adjusted so that the classification result matched the value that was obtained in the data set.

During this stage, many adjustments had to be made to the model's rules, as there were some parameters that did not correctly fit into any of the classification groups. However, the results were classified into the corresponding groups after the model was adjusted and all of the parameters were set right.

Tests

Once the training phase was completed, it was time to test the model and see if it was really trained to classify the data within the expected group. To perform these tests, we used the remaining 30 percent of the data, taking the resulting vector of each within the data set and introducing it as an input vector. The result was recorded for later analyzing all of the data and obtaining the effectiveness of the model once the process was completed.

These tests were performed with each of the different amplitudes and output groups that were mentioned above. We recorded the output value of the model in a table to later build a confusion matrix. We obtained six different configurations: amplitudes 0.3, 0.6, and 1 in combination with the three- and five-output groups.

The output value of each vector was concentrated in a table, categorizing it into three different groups depending on the result: correct, false positive, and false negative. Once the registration was finished, we applied the confusion matrix to obtain the percentage of the effectiveness of each configuration.

5.1. Results

Although we did not know whether the results would be promising at the beginning, we obtained an acceptable percentage of effectiveness (considering that no similar model had been done and we had to devise the variables based on the behavior of the criminals that were observed in the videos). After performing the tests and concentrating the results, we obtained the results presented in Table 2.

If we look at Table 2, we can see that the most effective results were the configurations for the three-group output with the medium amplitude; for the five-group output, the most effective was the low amplitude.

Table 2
Results of model

Groups	Amplitude	Accuracy [%]
Groups of three	0.3	86.21
	0.6	**90.73**
	1	85.42
Groups of five	0.3	**94.23**
	0.6	93.13
	1	85.42

It should be mentioned that the results will vary in each scenario; that is, if the implementation of the system is in a controlled scenario, the results will probably be better in each configuration. Similarly, if the scenario is less noisy, the voices have better quality, and both words and emotion are clearly identified, the results will be exceptional.

These are the corresponding results according to the groups of three with all of the amplitude tests. For the Tables 3–5, FN = False Negative and FP = False Positive (added accuracy column in Tables 3–5).

Table 3
Results for groups of three – amplitude = 0.3

	POS	**NEG**	**ACC**
POS	85	15 (FN)	86.21%
NEG	5 (FP)	60	

Table 4
Results for groups of three – amplitude = 0.6

	POS	**NEG**	**ACC**
POS	85	14 (FN)	90.73%
NEG	0 (FP)	66	

Table 5
Results for groups of three – amplitude = 1

	POS	**NEG**	**ACC**
POS	74	15 (FN)	85.42%
NEG	6 (FP)	70	

And now, these are the accuracy tables (Tabs. 6–8) for the tests on the groups of three with all of the amplitude tests.

Table 6

Accuracy results for Table 3

CORRECT	145
FALSE	20
ACCURACY	86.21%

Table 7

Accuracy results for Table 4

CORRECT	151
FALSE	14
ACCURACY	90.73%

Table 8

Accuracy results for Table 5

CORRECT	144
FALSE	21
ACCURACY	85.42%

These are the corresponding results according to the groups of five with all of the amplitude tests (for the Tables 9– 11, FN = False Negative and FP = False Positive).

Table 9

Results for groups of five – amplitude = 0.3

	POS	**NEG**	**ACC**
POS	85	4 (FN)	94.23%
NEG	5 (FP)	71	

Table 10

Results for groups of five – amplitude = 0.6

	POS	**NEG**	**ACC**
POS	87	4 (FN)	93.13%
NEG	7 (FP)	73	

Table 11

Results for groups of five – amplitude = 1

	POS	**NEG**	**ACC**
POS	74	16 (FN)	85.42%
NEG	5 (FP)	70	

And now, we present the accuracy results (Tabs. 12–14) for the amplitude variation on three groups; these are the accuracy tables for the tests on the groups of five with all of the amplitude tests.

Table 12
Accuracy results for Table 9

CORRECT	156
FALSE	9
ACCURACY	94.23%

Table 13
Accuracy results for Table 10

CORRECT	160
FALSE	11
ACCURACY	93.13%

Table 14
Accuracy results for Table 11

CORRECT	144
FALSE	21
ACCURACY	85.42%

6. Discussion and future work

Throughout this work, we have faced a great challenge since there are very few works in which a fuzzy system that is dynamic has been used that additionally takes context into account in order to adjust the values according to a user's needs. Although there are already contributions in this respect, none of them are completely dynamic (as is the objective of our work).

This research is part of a project that we are currently developing. We are in the first of three parts, since our intention is to offer a model that can be connected to various devices – especially security cameras that are located in public transport units in the cities with the highest incidences of crime in the country. We intend that this device will be developed in Arduino to carry out the first tests in a real way and, later, develop a device with our own design; first, the project must be completely finished.

The final project will have three stages (see Fig. 17):

- **Dynamic Fuzzy Model** (currently, the purpose of this paper);
- **Natural Language Processing** (we intend to use AI to analyze words that are transcribed to text and detect violent speech);
- **Sentiment Analysis** (with the use of AI, we will try to analyze sound in order to detect negative emotions such as anger or fury).

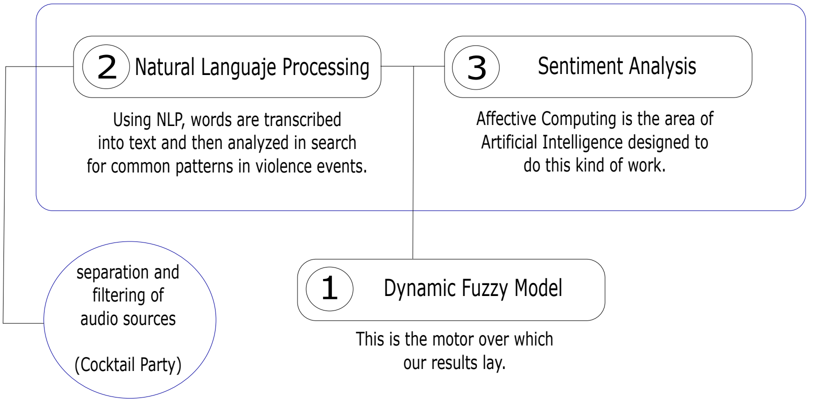

Figure 17. General project on which we are working

With our contribution, we intend to help public security agencies and, thus, reduce the crime rate in the country so that people and families can enjoy safer lives and cease being victims of crimes and assaults. INEGI has released a document that shows statistics about the perception of insecurity in each of the states of the republic.

This report shows (in percentages) how safe people feel in their homes, streets, public transportation, workplaces, and the neighborhoods in which they live. This work arises as a necessity to combat the violence that is experienced in almost all of the Mexican territory and that is a problem that many people and public organizations have wanted to solve but no one has succeeded thus far.

With our contribution, we do not intend to solve the problem since, for this, we need the collaboration of many organizations and security elements. Our main interest is to provide a tool that helps security systems detect these types of situations and alert the authorities automatically so that they can evaluate the situations and determine what actions to take.

7. Conclusions

This work is a great contribution because we have obtained very good results (even though it has not been implemented one hundred percent). According to the model that we have designed and the real data that we have been able to collect thanks to the publications that have been made through social networks and reliable platforms such as YouTube, good results are possible.

As mentioned earlier, this work is part of a three-stage project. Relative to the current project, we know that we can obtain great results in its implementation. The dynamic model that we propose is a novel model since it implements the dynamic part in each of its parts; with this, we obtain a versatile model that adapts to the variable changes of the needs of a scenario.

Given the good results that we obtained and the model variables that were designed according to real assault data, we can say that the model is reliable and will give accurate results when the project is completed. The dynamism of the model is a contribution to science that, to our knowledge, has not been made as of yet. However, the remaining stages still need to be developed for the contribution to be complete.

References

[1] Ahmed M.M., Isa N.A.M.: Information granularity model for evolving context-based fuzzy system, *Applied Soft Computing*, vol. 33, pp. 183–196, 2015.

[2] Baker M.T., Van Hasselt V.B., Sellers A.H.: Validation of the Novaco Anger Scale in an Incarcerated Offender Population, *Criminal Justice and Behavior*, vol. 35(6), pp. 741–754, 2008.

[3] Bautista-Duràn M., García-Gómez J., Gil-Pita R., Sánchez-Hevia H., Mohino-Herranz I., Rosa-Zurera M.: Acoustic Detection of Violence in Real and Fictional Environments. In: *ICPRAM*, pp. 456–462, 2017.

[4] Bhaskar J., Sruthi K., Nedungadi P.: Hybrid approach for emotion classification of audio conversation based on text and speech mining, *Procedia Computer Science*, vol. 46, pp. 635–643, 2015.

[5] Bugueño Sáez V.G.: Modelo de detección de agresiones verbales, por medio de algoritmos de Machine Learning, 2017, Repositorio Académico of Universidad de Chile. https://repositorio.uchile.cl/handle/2250/148580?

[6] Butko T.: *Feature selection for multimodal: acoustic Event detection*, Universitat Politècnica de Catalunya, 2011.

[7] Carletti V., Foggia P., Percannella G., Saggese A., Strisciuglio N., Vento M.: Audio surveillance using a bag of aural words classifier. In: *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 81–86, IEEE, 2013.

[8] Chang J.C., Wang P.S., Fan K.H., Yang S.R., Su D.Y., Lin M.S., Sun M.T., Tseng Y.C.: iMace: protecting females from sexual and violent offenders in a community via smartphones. In: *2011 40th International Conference on Parallel Processing Workshops*, pp. 71–74, IEEE, 2011.

[9] Cheng M., Cai K., Li M.: RWF-2000: An open large scale video database for violence detection. In: *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 4183–4190, IEEE, 2021.

[10] Giannakopoulos T.: Study and application of acoustic information for the detection of harmful content, and fusion with visual information, *Department of Informatics and Telecommunications, vol PhD University of Athens, Greece*, 2009.

[11] Giannakopoulos T., Kosmopoulos D., Aristidou A., Theodoridis S.: Violence content classification using audio features. In: *Hellenic Conference on Artificial Intelligence*, pp. 502–507, Springer, 2006.

[12] Ho D.T., Garibaldi J.M.: Context-dependent fuzzy systems with application to time-series prediction, *IEEE Transactions on Fuzzy Systems*, vol. 22(4), pp. 778–790, 2013.

[13] Iancu I.: A Mamdani type fuzzy logic controller, *Fuzzy logic – controls, concepts, theories and applications*, pp. 325–350, 2012.

[14] Islam N., Hossain M.R., Anisuzzaman M., Obaidullah A.J.M., Islam S.S.: Design and Implementation of Women Auspice System by Utilizing GPS and GSM. In: *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 1–6, IEEE, 2019.

[15] Maerker D.: Choferes del Edomex, hartos de los robos a transporte, 2022. https://noticieros.televisa.com/videos/choferes-del-edomex-hartos-de-los-robos-a-transporte/. Last accessed 29 March 2022.

[16] Mills J.F., Kroner D.G., Forth A.E.: Novaco Anger Scale: Reliability and validity within an adult criminal sample, *Assessment*, vol. 5(3), pp. 237–248, 1998.

[17] Naik N., Diao R., Shen Q.: Dynamic fuzzy rule interpolation and its application to intrusion detection, *IEEE Transactions on Fuzzy Systems*, vol. 26(4), pp. 1878–1892, 2017.

[18] Nakamura A.: De paseo en la CDMX? Cuidado, éstas son las zonas más peligrosas, 2017. https://www.nacion321.com/seguridad/de-paseo-en-la-cdmx-cuidado-estas-son-las-zonas-mas-peligrosas. Last accessed 27 December 2021.

[19] Pancardo P., Hernández-Nolasco J.A., Wister M.A., Garcia-Constantino M.: Dynamic Membership Functions for Context-Based Fuzzy Systems, *IEEE Access*, vol. 9, pp. 29665–29676, 2021.

[20] Park H.S., Yoo J.O., Cho S.B.: A context-aware music recommendation system using fuzzy bayesian networks with utility theory. In: *International Conference on Fuzzy Systems and Knowledge Discovery*, pp. 970–979, Springer, 2006.

[21] Penet C., Demarty C.H., Gravier G., Gros P.: Audio event detection in movies using multiple audio words and contextual Bayesian networks. In: *2013 11th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pp. 17–22, IEEE, 2013.

[22] Periodico Oficial, Gobierno del Estado Libre y Soberano de México, 2018. http://legislacion.edomex.gob.mx/sites/legislacion.edomex.gob.mx/files/files/pdf/gct/2018/mar204.pdf. Last accessed 20 December 2021.

[23] Pokorny F.B., Graf F., Pernkopf F., Schuller B.W.: Detection of negative emotions in speech signals using bags-of-audio-words. In: *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 879–884, IEEE, 2015.

[24] Pradeep M., Abinaya R., Anandhi S.S., Soundarya S.: Intelligent Safety System to Prevent Acid Attacks, *Asian Journal of Applied Science and Technology (AJAST)*, vol. 1(3), pp. 243–248, 2017.

[25] Ramli N., Mohamad D.: A comparative analysis of centroid methods in ranking fuzzy numbers, *European Journal of Scientific Research*, vol. 28(3), pp. 492–501, 2009.

[26] Robo a Uber en Ecatepec no fue hecho por usuarios, 2018, Periódico Excelsior. https://www.excelsior.com.mx/comunidad/robo-a-uber-en-ecatepec-no-fue-hecho-por-usuarios/1261100. Last accessed 11 May 2022.

[27] Rodellar-Biarge V., Palacios-Alonso D., Nieto-Lluis V., Gómez-Vilda P.: Towards the search of detection in speech-relevant features for stress, *Expert Systems*, vol. 32(6), pp. 710–718, 2015.

[28] Rouas J.L., Louradour J., Ambellouis S.: Audio events detection in public transport vehicle. In: *2006 IEEE Intelligent Transportation Systems Conference*, pp. 733–738, IEEE, 2006.

[29] Santos F., Durães D., Marcondes F.S., Hammerschmidt N., Lange S., Machado J., Novais P.: In-Car Violence Detection Based on the Audio Signal. In: *International Conference on Intelligent Data Engineering and Automated Learning*, pp. 437–445, Springer, 2021.

[30] Shulby C., Pombal L., Jordão V., Ziolle G., Martho B., Postal A., Prochnow T.: Proactive Security: Embedded AI Solution for Violent and Abusive Speech Recognition. In: *2018 7th Brazilian Conference on Intelligent Systems (BRACIS)*, pp. 248–253, IEEE, 2018.

[31] Siantikos G., Sgouropoulos D., Giannakopoulos T., Spyrou E.: Fusing multiple audio sensors for acoustic event detection. In: *2015 9th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pp. 265–269, IEEE, 2015.

[32] Souza-Leal B., De-Carvalho C.A., Antunes E.: La violencia contra mujeres brasileñas en las esferas pública y mediática, *Comunicar*, vol. 55, pp. 19–28, 2018. doi: 10.3916/C55-2018-02.

[33] Tasa de incidencia delictiva por entidad federativa de ocurrencia por cada cien mil habitantes, Instituto Nacional de Estadística y Geografía, 2019. https://www.inegi.org.mx/temas/incidencia/. Last accessed 22 December 2021.

[34] Taskin A., Kumbasar T.: An open source Matlab/Simulink toolbox for interval type-2 fuzzy logic systems. In: *2015 IEEE Symposium Series on Computational Intelligence*, pp. 1561–1568, IEEE, 2015.

[35] Timmons A.C., Chaspari T., Han S.C., Perrone L., Narayanan S.S., Margolin G.: Using Multimodal Wearable Technology to Detect Conflict among Couples, *Computer*, vol. 50(3), pp. 50–59, 2017.

[36] Zhao J., Bose B.K.: Evaluation of membership functions for fuzzy logic controlled induction motor drive. In: *IEEE 2002 28th Annual Conference of the Industrial Electronics Society. IECON 02*, vol. 1, pp. 229–234, IEEE, 2002.

[37] Zhuang X., Zhou X., Hasegawa-Johnson M.A., Huang T.S.: Real-world acoustic event detection, *Pattern Recognition Letters*, vol. 31(12), pp. 1543–1551, 2010.

[38] Zieger C., Brutti A., Svaizer P.: Acoustic based surveillance system for intrusion detection. In: *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 314–319, IEEE, 2009.

Affiliations

Obed Campos
Autonomous Juarez University of Tabasco, Academic Division of Science & Information Technologies, 191h18002@egresados.ujat.mx

Pablo Pancardo
Autonomous Juarez University of Tabasco, Academic Division of Science & Information Technologies, pablo.pancardo@ujat.mx

José Adán Hernández-Nolasco
Autonomous Juarez University of Tabasco, Academic Division of Science & Information Technologies, adan.hernandez@ujat.mx

Received: 15.01.2022
Revised: 24.07.2022
Accepted: 27.08.2022