Revista Ciência Agronômica, v. 52, n. 4, e20196877, 2021
Centro de Ciências Agrárias - Universidade Federal do Ceará, Fortaleza, CE
www.ccarevista.ufc.br

Scientific Article
ISSN 1806-6690

Neural network and clustering techniques for tractor accidents on highways in the south-east of Brazil[1]

Rede neural e técnicas de agrupamento em acidentes com tratores em rodovias na Região Sudeste

Deivielison Ximenes Siqueira Macedo[2]*, Viviane Castro dos Santos[2], Leonardo de Almeida Monteiro[2], Jefferson Auteliano Carvalho Dutra[2], José Wally Mendonça Menezes[3]

ABSTRACT - Until recently, accident indicators were analysed separately due to the methods employed, however, the joint use of neural networks and clustering techniques has proven to be an excellent tool for analysing how accidents occur. As such, the aim of this study was to use neural networks and cluster analysis on accident indicators involving tractors on federal highways in the south-east of Brazil. A total of 496 incidents were analysed between 2007 and 2016. The indicators for the accidents under evaluation were time, type of accident, cause of accident, weather conditions, condition of the victims, road layout and federated state. The use of neural networks was based on self-organising maps (SOM), hierarchical clustering employing dendrograms, and non-hierarchical clustering employing the k-means coefficient. Using these techniques, it was possible to divide the incidents into 18 accident groups, of which 11 were represented by the state of Minas Gerais, one group where casualties were predominant, and one group with fatalities. It proved possible to analyse the factors that led to the accident, together with its consequences. Machine traffic during periods of low natural light on straight roads caused rear-end collisions, with casualties and fatalities.

Key words: SOM networks. k-means. Safety, Incidents. Agricultural machinery.

RESUMO - Anteriormente os acidentes tinham seus indicadores analisados separadamente devido os métodos utilizados, todavia o uso em conjunto de redes neurais e técnicas de agrupamento tem se mostrado uma excelente ferramenta de análise de situações de ocorrência de sinistros. Assim objetivou-se realizar o uso de redes neurais e análise de agrupamento sobre os indicadores dos acidentes com tratores nas rodovias federais na região Sudeste. Foram analisadas 496 ocorrências entre o período de 2007 a 2016. Os indicadores dos sinistros avaliados foram: horário, tipo de acidente, causa do acidente, condições climáticas, condições dos acidentados, traçado da via e unidade federativa. O uso das redes neurais se deu pelos mapas auto-organizados-SOM, os métodos de agrupamento hierárquico por dendrograma e o não hierárquico pelo coeficiente de k-means. Através das técnicas foi possível dividir as ocorrências em 18 grupos de acidentes, dos quais 11 foram representados pelo estado de Minas Gerais, 1 grupo com dominância de vítimas feridas e 1 grupo com vítimas fatais. Foi possível analisar os fatores em conjunto que levaram a ocorrência dos sinistros e a consequência do mesmo. O tráfego de máquinas em períodos com pouca luminosidade natural em pistas retas, ocasionaram colisões traseiras com vítimas feridas e vítimas fatais.

Palavras-chave: Redes SOM. K-means. Segurança. Sinistros. Máquinas agrícolas.

DOI: 10.5935/1806-6690.20210058
Editor-in-Chief: Prof. Alek Sandro Dutra - alekdutra@ufc.br
*Author for correspondence
Received for publication 20/08/2019; approved on 07/07/2021

[1]Parte da Tese do primeiro autor, apresentado ao Programa de Pós-Graduação em Engenharia Agrícola, Centro de Ciências Agrárias (CCA), Universidade Federal do Ceará (UFC)

[2]Departamento de Engenharia Agrícola, Centro de Ciências Agrárias (CCA), Universidade Federal do Ceará (UFC), Programa de Pós Graduação em Engenharia Agrícola (DENA), Fortaleza-CE, Brasil, derilsiqueira@hotmail.com (ORCID ID 0000-0003-1860-9212), viviane.castro@ufc.br (ORCID ID 0000-0002-3688-8043), aiveca@ufc.br (ORCID ID 0000-0003-3703-8072), jeffersocarvalho@hotmail.com (ORCID ID 0000-0001-5314-9077)

[3]Departamento de Telemática, Programa de Pós Graduação em Engenharia em Telecomunicações (PPGET), Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE), Fortaleza-CE, Brasil, wally@ifce.edu.br (ORCID ID 0000-0003-4819-2630)

INTRODUCTION

Accidents with agricultural machinery remain constant, both on and off farming properties. Macedo *et al.* (2015a) point out that on public roads in Minas Gerais, for example, the problem is growing; this is because the state has the largest road network in the country, where the speed of passenger vehicles is high compared to that of tractors.

Bellochio *et al*. (2018) report the importance of studying accidents with tractors in the south-east. According to them, this region, together with the south and mid-west, are responsible for more than 50% of incidents on public roads; the authors also report that the importance of agribusiness in the country promotes the use of these machines, and the reality of their circulating on public roads is difficult to reverse, requiring action if accidents are to be avoided.

Despite the number of cases, there was a certain difficulty in analysing this information due to the restrictions of the techniques used for qualitative data, generally frequency analysis, which allows only one accident class to be analysed at a time; however, according to Macedo *et al*. (2018), the use of neural networks by means of self-organising maps and clustering techniques makes it possible to check patterns and correlations between indicators, in addition to profiling and obtaining a more accurate view of how these incidents occur.

According to Kohonen (2013), the self-organising map, or SOM network, is a neural network that works as an automatic method of data analysis. It represents the distribution of input data using a finite set of models, which are associated with nodes on a two-dimensional grid, generally so that more-similar models automatically combine at adjacent nodes and more-dissimilar models move apart. It is therefore possible to obtain a diagram of the similarity between models and observe the topographic relationships of the information, especially in a large set of data. Hu *et al*. (2019) report that the SOM network is a powerful visualisation tool for creating two-dimensional projections, which, however, remain discrete.

K-means clustering is a non-hierarchical agglomerative method that attempts to minimise the sum of the Euclidean distance and the mean square as a function, efficiently generating well-defined clusters or groups (ISMKHAN, 2018); however it requires a base number of groups when programming that can be determined by the dendrogram (MACEDO *et al*., 2018).

Comberti, Demichela and Baldissone (2018) report that combining self-organising maps with the k-means coefficient is an efficient tool for analysing data related to occupational accidents due to the ability to clearly group and visualise the data, and helps the user make decisions to prevent accidents.

According to Palamara, Pglione and Piccinini (2011), the joint use of these techniques is a powerful method of retrieving information from huge databases and allows any information hidden in the dataset to be found.

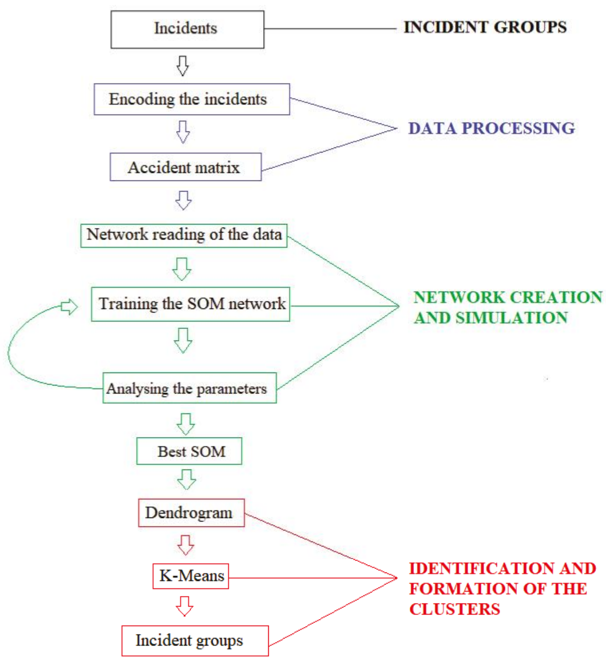
Analysing the occurrence of accidents with tractors is still seldom used, little explored and, in most cases, employs the use of restrictive tools. As such, the aim of this study was to use neural networks and clustering techniques with information of incidents involving tractors on federal highways in the south-east of the country, to identify any patterns, determine the accident groups that are most similar to each other, and verify the sequence of classes that resulted in greater risk to the injured.

MATERIAL AND METHODS

The present study was conducted at the **Laboratory** for **Agricultural Machine Accident Investigation –** LIMA. It consists in the use of artificial neural networks by means of self-organising maps, and clustering techniques employing dendrograms and the k-means coefficient, for accidents involving tractors on federal highways in the south-east of Brazil (Figure 1) from 2007 to 2016, with a total of 496 incidents being analysed.

Database

Figure 1 - Flowchart showing the structure of the procedures carried out



The data were obtained from the Federal Highway Police (FHP) in Fortaleza. The information was collected from Traffic Accident Bulletins (TAB). These bulletins include the various accident indicators with their respective classes of occurrence.

An occurrence registered in a TAB has several indicators and these indicators are characterised by certain classes. In this study, seven occurrence indicators were used, namely: type of accident (14 classes), cause of accident (12 classes), time of occurrence (8 classes), weather conditions (6 classes), clinical situation of the accident victims (5 classes), road layout (3 classes) and federated state (4 classes).

Processing the data

Information regarding accidents is qualitative; however, to better represent and process the information, the data were coded to give whole numbers, where each class was assigned a number, taking into account the possible degree of approximation between classes as a determination criterion (Table 1).

Network creation and simulation

The neural network used was the Kohonen network (SOM network), or self-organising maps. This is an unsupervised network, for which the output result is undetermined. It is competitive, where neurons compete to define the winning neuron for each interaction, and as such, produces good results in which each winning neuron indicates a pattern of data behaviour.

Table 1 - Transformation of the indicator classes into numerical values

TYPE							
CLASS	Striking the victim	Potential damage	Spilled load	Fire	Fall	Leaving the road	Overturned vehicle
CODE	1	2	3	4	5	6	7
CLASS	Collision with bicycle	Collision with fixed object	Collision with movable object	Transversal collision	Rear-end collision	Head-on collision	Side collision
CODE	8	9	10	11	12	13	14
CAUSE							
CLASS	Animals on the road	Defect in the road	Mechanical defect of the vehicle	Disobeying signs	Not maintaining a safe distance	Wrongful overtaking	
CODE	1	2	3	4	5	6	
CLASS	Undue speed	Lack of attention	Sleeping	Alcohol consumption	Other		
CODE	7	8	9	10	11		
TIME OF DAY							
CLASS	Small hours (00:00 to 02:59)	Early hours (03:00 to 05:59)	Start of day (06:00 to 08:59)	End of morning (09:00 to 11:59)			
CODE	1	2	3	4			
CLASS	Start of afternoon (12:00 to 14:59)	End of afternoon (15: to 17:59)	Start of evening (18:00 to 20:59)	End of night (21:00 to 23:59)			
CODE	5	6	7	8			
WEATHER CONDITIONS							
CLASS	Sunny/Clear	Fog/Mist	Wind/Rain	Cloudy	Hail		
CODE	1	2	3	4	5		
CLINICAL CONDITION OF THE VICTIMS							
CLASS	No casualties	Casualties	Bad casualties	One fatality and casualties	Fatalities		
CODE	1	2	3	4	5		
ROAD LAYOUT							
CLASS	Straight	Curved	Intersection				
CODE	1	2	3				
STATE							
CLASS	Minas Gerais	São Paulo	Rio de Janeiro	Espírito Santo			
CODE	1	2	3	4			

During the process of creating and simulating the network, Haykin (2001) reports that the basic processes of the network include initialisation, sampling, similarity matching (Equation 1), updating (Equation 2) and continuation. During initialisation, the initial weight vectors are determined; during sampling, sample x is removed by the system; during similarity, the winning neuron is determined; during update, the synaptic weight is updated, and the learning rate and neighbourhood function of the winning neuron are changed; finally, during continuation, the process is repeated until no significant changes are seen in the map.

$$i(x) = argmim_j \left\| x(n) - w_j \right\| \qquad (1)$$

where: w_j = the ith weight vector; $x(n)$ = vector; $i(x)$ = winning neuron.

$$w_j(n+1) = w_j(n) + n(n)h_{j,i(x)}(n)(x(n) - w_j(n)) \qquad (2)$$

where: $n(n)$ = learning rate parameter; $h_{j,i}(n)$ = neighbourhood function centred around the winning neuron $i(x)$.

When generating the maps, their representativeness and quality were evaluated. This evaluation was carried out based on the quantisation error (Equation 3), the topographic error (Equation 4), and the size of the map in relation to the number of occurrences. The quantisation error shows whether the input vectors are well adjusted, while the topographic error is the proportion of discontinuous objects on the map, where the closer the values are to zero, the better the representation.

$$Te = \frac{1}{n}\sum_{k=1}^{n} \mu(x_k) \qquad (3)$$

where: Te = topographic error; n = total number of patterns; $u(x_k)$ = 1 if the first and second winning neurons are not adjacent; $u(x_k)$ = 0 if the first and second neurons are neighbours.

$$Qe = \frac{\sum_{k=1}^{n} \left\| x_k - \omega_{BMU} \right\|}{n} \qquad (4)$$

where: Qe = quantisation error; x_k = input vector; ω_{BMU} = weight vector.

The networks were simulated numerous times, taking into account the errors and the configuration of neurons. According to Kohonen (2013), dimensioning a SOM network is a method of trial and error that requires various training sessions with different sizes, taking into account a sufficiently good resolution and statistical accuracy, and avoiding oversizing. Whenever the values were unsatisfactory, the training was repeated with a different configuration of neurons until the best configuration was decided. This was then followed by the next step.

Identification and formation of clusters

The U-matrix, the map generated by the Kohonen network, is not always clear in relation to its delimitations, so it is interesting to use this technique together with such clustering techniques as the agglomerative dendrogram (Equation 5) to determine the number of groups and the k-means coefficient (Equation 6) for the actual division of these groups in the matrix. The dendrogram uses similarity measurements, such as the Euclidean distance, to determine the number of groups, while the k-means coefficient divides a set of similar data and classes, also using similarity measurements.

$$d_{if} = \left[\sum_{j=1}^{p} \left(X_{ij} - X_{f_j}\right)^2\right]^{\frac{1}{2}} \qquad (5)$$

$$C_i = \frac{1}{m_i}\sum_{j=1}^{m_i} X_{ji} \qquad (6)$$

where: *ci* = centroid of cluster *Ci*; *mi* = mi = number of xj data grouped in cluster *Ci.*

After generating the group maps, they were analysed and separated, taking into account all the indicators under evaluation. Each group was represented by the most frequent class within that indicator.

Creation and simulation of the network along with the formation of the clusters was carried out using the SOM Toolbox of the Matlab 2010 software, while the separation, tabulation and division of the occurrences was carried out in Excel.

RESULTS AND DISCUSSION

More than 15 training sessions were held, each with a different configuration. The one that presented the best results, considering the topographic error (0.02) and quantisation error (0.893), was the 20x20 configuration, with a total of 400 neurons.

The result of the training can be seen in Figure 2. Evaluating the plane for the federated state shows a large number of active neurons for Minas Gerais (code 1), very few active neurons for São Paulo (code 2), and a greater quantity for Rio de Janeiro (code 3) and Espírito Santo (code 4), albeit less than for Minas Gerais.

Olawoyin *et al.* (2013) reports that the results generated when training the network can easily be visualised from the component planes, where each hexagon is equal to one unit that is found in the same location in the U-matrix.

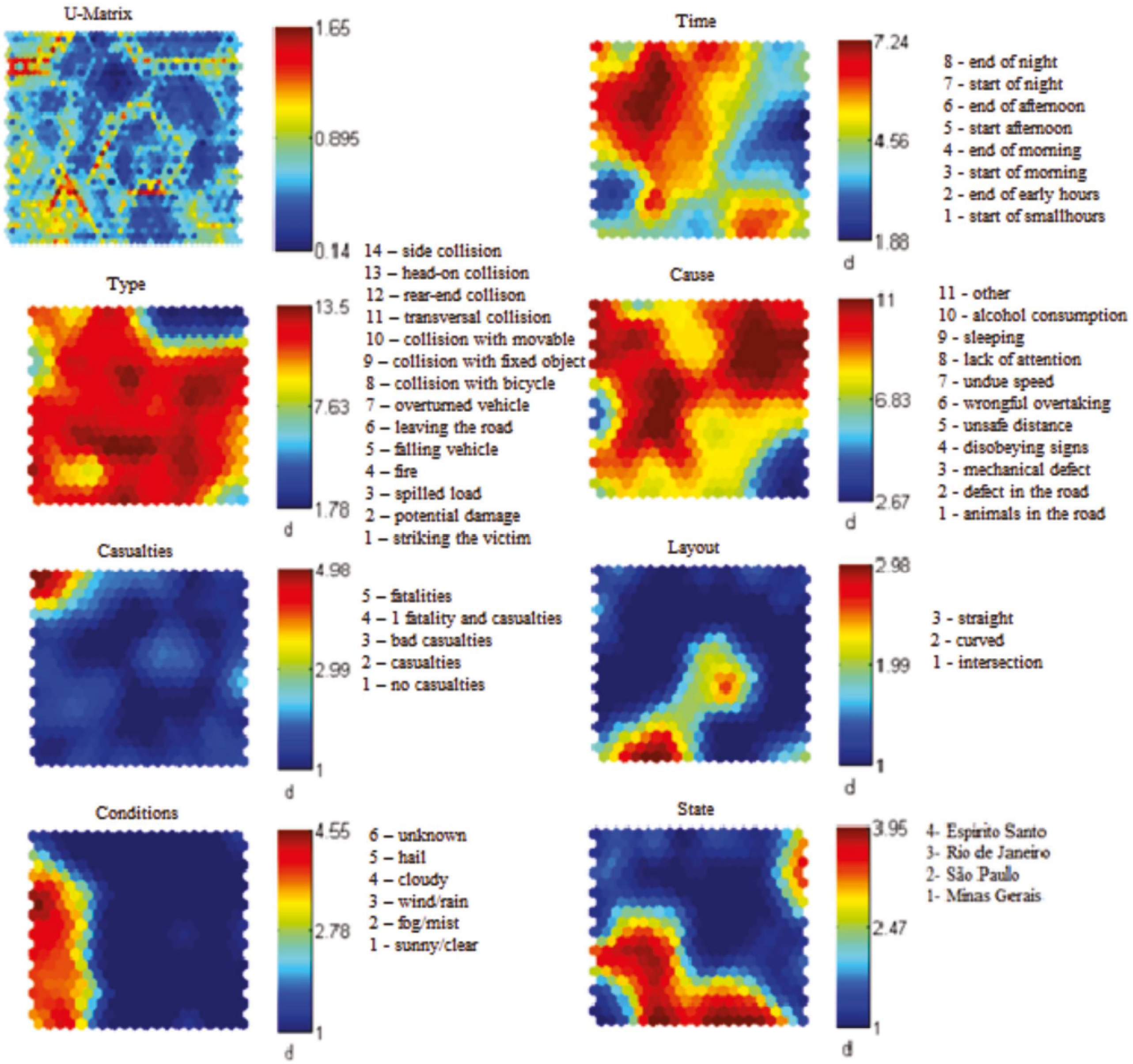
For the type of accident, the ‘rear end collision’ class (code 12) is the most frequent, whereas for the condition of the victims it is the class with no casualties (code 1). The weather-conditions class with the most active neurons was the ‘sunny/clear’ class (code 1). Finally the most frequent class of road layout was the ‘straight road’ class (code 1). Although ‘type of accident’ apparently has more neurons activated by rear-end collision, the plane does not show very distinct class zones, as do the other indicators; as a result, this distribution can affect the distribution of the groups.

The cause indicator has two dominant classes with more active neurons, they are ‘lack of attention’ (code 8) and ‘other causes’ (code 11). The time-of-day indicator had several neurons activated by different codes, and did not have a dominant colour, thereby making it impossible to assert a time class that better represents the indicator. If ‘time of day’ had been divided into larger classes, one class might well have been dominant, but this could have masked the actual time.

In their study, Macedo *et al*. (2015b) obtained ‘rear-end collision’ and ‘lack of attention’ as the main classes of occurrence for their indicators, with the authors recommending taking regular breaks during the working period and transporting the machine on the roads only when really necessary, as per the current traffic legislation.

The generated U-matrix determined an average distance of 0.895; the lower this value, the greater the rigour in selecting the groups. The topological representation can be seen in Figure 3. Ji-Hong, Jian-Cheng and Nan (2011) state that the U-matrix, or the unified distance matrix, is the most usual way of representing training data.

Figure 2 – U-matrix and component planes for accidents in the south-east



When viewing the matrix, neurons with smaller distances have colder colours and are in the depressions, whereas more-distant neurons have warmer colours and are found on the slopes. Using this view, it is not possible to verify the number of groups with certainty, and to validate the number of groups, a dendrogram is necessary (Figure 4).

The horizontal cut-off line of the dendrogram is at the height of the mean distance obtained with the U-matrix (0.895), from which it is possible to determine 18 groups. The k-means coefficient was later used to generate the boundaries of the groups, which were projected onto the U-matrix of the accidents (Figure 5) so as to visualise the groups.

The groups on the left were allocated this position as they are more distant from the other groups. This is due to the class of incidents in these groups: for example, Group 2 is in the zone allocated by the network to accidents with fatalities and serious casualties, differing from the groups on the right, which mostly had accidents with no casualties, as can be seen in the component plane for the 'clinical conditions of the victims' indicator.

Another indicator that allocated these groups to the same area was related to weather conditions, where accidents with different classes than 'sunny/clear' were positioned to the left of the map. Finally, the road layout indicator also contributes to this separation and increase in distance, moving 'curved roads' and 'intersections' from the centre to the left, despite also showing isolated zones.

Table 2 shows the predominant groups in Espírito Santo and Rio de Janeiro. Of the states in the south-east, the only state not represented by a group was São Paulo. Few neurons were activated by accidents in the state, indicating a low number of incidents with tractors on the federal highways compared to other states in the region.

In most groups, the 'time of day' indicator had more than one class representing the group, this situation also occurred with 'type of accident', but in smaller proportions, and may have been due to the proximity of the classification used in encoding and the excessive number of occurrences of the same class for different indicators: for example, 'straight road' for 'road layout', or even good weather conditions.

The association of factors seen in Group 1 demonstrates that even under conditions of good visibility the greater flow of vehicles, especially at the start or end of the working day, can cause congestion on the roads due to the low speed of the tractor. In such situations, probably due to being on a straight road under good conditions, the drivers tried to overtake the tractor, but due to a lack of attention or knowledge of the characteristics of the tractor, they collided with the side or even with the rear of the machine.

Figure 3 - Topological matrix of accidents with tractors on federal highways in the south-east

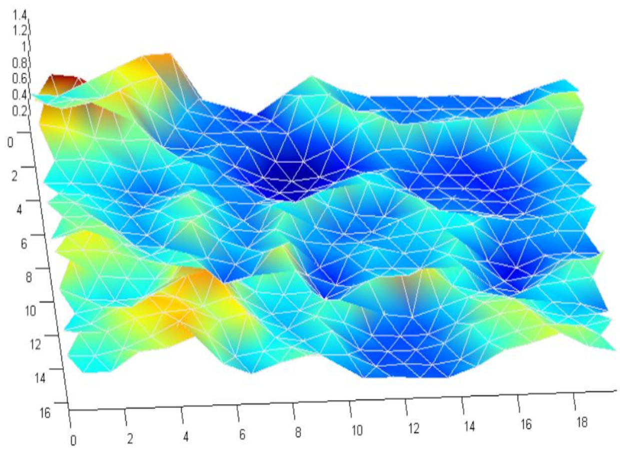


Figure 4 - Dendrogram of accidents with tractors on federal highways in the south-east

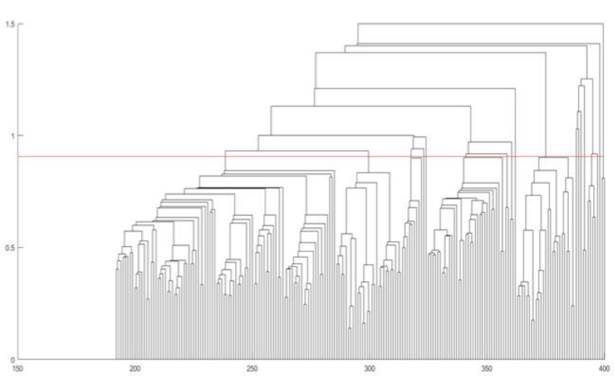


Figure 5 – U-matrix with the group projections obtained using the k-means coefficient

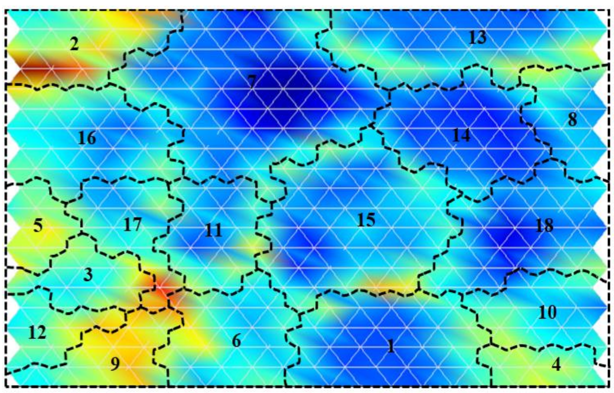


Table 2 - Breakdown of accident groups in Espírito Santo and Rio de Janeiro

INDICATOR		GROUP						
		1	3	4	6	8	11	17
Number of accidents		42	13	13	22	25	15	14
State	Class	Espírito Santo	Espírito Santo and Rio de Janeiro	Espírito Santo	Espírito Santo and Rio de Janeiro	Rio de Janeiro	Rio de Janeiro	Rio de Janeiro
	Freq. (%)	57.14	76.92	46.15	100.00	52.00	46.67	57.14
Time of day	Class	Start and End of Afternoon	Early Hours and Start of Day	End of Morning and Start of Afternoon	End of Morning and Start of Afternoon	Start of Day and End of Morning	End of Afternoon	End of Afternoon and Start of Evening
	Freq. (%)	54.76	61.54	53.84	59.09	68.00	53.33	85.71
Type	Class	Rear-end and Side Collision	Rear-end Collison	Overturning and Leaving the Road	Side Collision	Rear-end and Side Collision	Rear-end and Side Collision	Rear-end and Side Collision
	Freq. (%)	71.42	46.15	61.15	50.00	60.00	86.66	64.28
Cause	Class	Lack of Attention	Other	Mechanical Defect of the Vehicle	Lack of Attention	Other	Other	Other
	Freq. (%)	66.67	46.15	76.92	50.00	80.00	100.00	64.29
Conditions	Class	Sunny/Clear	Cloudy	Sunny/Clear	Sunny/Clear	Sunny/Clear	Sunny/Clear	Cloudy
	Freq. (%)	100	69.23	100	100	100	100	57.14
Casualties	Class	No Casualties	Casualties	No Casualties	No Casualties	No Casualties	No Casualties	No Casualties
	Freq. (%)	78.57	69.23	61.54	72.73	68.00	60.00	71.43
Road	Class	Straight	Straight	Curved	Curved	Straight	Straight	Straight
	Freq.(%)	100	100	76.92	54.55	96	86.67	100

Olson, Soccolich and Hanowski (2019) state that distraction leads to a delay in reacting to information, and is responsible for most incidents involving heavy vehicles. In most cases, this lack of attention results in a collision between those involved.

Group 3 was the only group with accidents during the small or early hours (00:00 to 05:59). This period is characterised by hours of darkness only illuminated by public lighting. In addition, the weather conditions at the time of the accidents were cloudy, which further reduced visibility. Under such conditions it is difficult for drivers to identify the tractor, and possibly for this reason, this group included people who were injured in the accidents. In their study, Wang *et al*. (2019) found that despite a smaller number of incidents, accidents occurring between 02:00 and 05:59 had more fatalities than those occurring at other times.

The accident profile verified in Group 4 has some sensitive points that need to be addressed. First are mechanical defects. Accidents due to mechanical defects can occur due to simple problems, such as the direction-indicator lights, to more serious problems, such as the tractor brakes. From the conditions of the profile, the tractor probably had problems with braking or steering. Accidents due to mechanical defects are extremely worrying, and the periodic maintenance of a tractor must be carried out correctly. Operators can obtain more detailed information on how and when to carry out maintenance during courses on the operation and maintenance of agricultural machinery. However, in many cases this is a distant reality. Schlosser *et al*. (2002) found in their study that around 61% of operators interviewed did not attend any course on operating a tractor.

For ‘road layout’, curved roads or intersections require greater skill and knowledge by users, not only to perform the manoeuvre, but also to visualise the surroundings, considering that vision may be impaired.

Overturning is common on agricultural properties, mainly due to the lack of uniform terrain, unlike the situation on a federal highway. As overturning is characterised by a lack of casualties, the tractors involved were probably new, less than 10 years old, with ROPS, and the operator was probably wearing a seat belt, otherwise they could be thrown out of the tractor as the machine turned on its axis. It can be assumed that the operators were returning with the tractor at the end of their working shift, and the machine presented a problem that prevented the operator from going around a curve, getting too close to the shoulder, to a point where the machine tipped over or left the track without causing major physical damage to the operator.

Preventing overturning is of paramount importance. As pointed out by Antunes, Cordeiro and Teixeira (2018) in their study carried out in Portugal, an overturned tractor is the most common type of accident, and is responsible for various types of injury, including fatalities.

Groups 8 and 11 were very similar regarding class, and differed only in relation to the time of the incident, both groups occurring due to other causes. Fernandes *et al*. (2014) found in their study that the second most frequent cause of accidents was tiredness (24.52%), second only to a lack of attention (26.31%). Accidents that occurred due to tiredness can be classified as due to 'other causes', since there are none found in the TAB. In addition to tiredness, several possibilities could be considered other causes, such as extreme conditions and loss of control, among others; this shows the range of possible incidents linked to the class of 'other causes'.

The incidents in group 17 possibly occurred due to the low natural light at the time of the incident and low visibility due to the weather.

The state of Minas Gerais had the largest number of groups in the study (Table 3), i.e. with a greater range of possible accidents involving tractors on federal highways, of sufficient size for the neural network to understand these groups as being distinct.

Group 2 was the only group in the south-east that involved fatalities, which differentiates it from other cases of rear-end collision in the region. Another important factor in this group is cause. The group includes the class 'other causes', and considering that other indicators did not contribute significantly, there is a wide range of possible causes for these fatalities. One of the possibilities in this group involves rear-end collisions between motorcycles and tractors, which can be more severe than rear-end collisions involving cars and tractors, even at low speeds. The combination of low natural lighting during the late afternoon and early evening, and straight roads, may have made the driver of the rear vehicle feel more secure; however, under such conditions, seeing the object in front can be difficult, not allowing sufficient time between seeing the obstacle, reacting and safely braking.

In their study, Zhang, Yau and Zhang (2014) found a greater trend towards serious injury or death at night under public lighting, with this trend increasing in the absence of lighting.

Table 3 - Breakdown of accident groups in Minas Gerais

INDICATOR		GROUP										
		2	5	7	9	10	12	13	14	15	16	18
Number of accidents		24	13	58	21	22	13	53	37	45	31	35
State	Class	Minas Gerais	Minas Gerais	Minas Gerais	Minas Gerais	Minas Gerais	Minas Gerais	Minas Gerais	Minas Gerais	Minas Gerais	Minas Gerais	Minas Gerais
	Freq. (%)	83.33	61.54	87.93	52.38	81.82	76.92	56.60	97.30	91.11	64.52	82.86
Time of day	Class	End of evening and start of night	End of morning Start of afternoon	End of afternoon and start of evening	End of morning Start of afternoon	End of morning Start of afternoon	Start and end of morning	End of morning and end of afternoon	End of morning Start of afternoon	Start and end of afternoon	Start and end of night	Start and end of morning
	Freq. (%)	54.16	61.53	70.68	61.90	59.09	69.23	62.26	70.27	60.00	58.06	77.14
Type	Class	Rear-end	Rear-end	Rear-end	Side and transversal	With fixed object	Side	Potential damage	Side and rear-end	Side	Rear-end and side	Rear-end and transversal
	Freq. (%)	45.83	46.15	37.93	47.61	31.82	46.15	66.04	64.86	42.22	41.93	62.86
Cause	Class	Other	Mech. defect	Lack of attention	Lack of attention	Mech. defect	Lack of attention	Other	Other	Lack of attention	Other	Lack of attention
	Freq. (%)	58.33	46.15	68.97	52.38	63.64	46.15	67.92	97.30	57.78	61.29	82.86
Conditions	Class	Sunny/ Clear	Cloudy	Sunny/ Clear	Cloudy	Sunny/ Clear	Cloudy	Sunny/Clear	Sunny/ Clear	Sunny/ Clear	Cloudy	Sunny/Clear
	Freq. (%)	83.33	76.92	98.28	61.90	100	84.62	100	100	97.78	48.39	94.29
Casualties	Class	Fatalities	No casualties	No casualties	No casualties	No casualties	No casualties	No casualties	No casualties	No casualties	No casualties	No casualties
	Freq. (%)	62.50	53.85	56.90	76.19	54.55	84.62	66.04	70.27	68.89	58.06	57.14
Road	Class	Straight	Straight	Straight	Curved	Straight	Straight	Straight	Straight	Curved	Curved	Straight
	Freq. (%)	75	100	98.28	61.90	95.45	61.54	84.91	100	86.67	83.87	100

The profile seen in Group 5 was probably due to failures related to the signalling system of the tractor, so when braking, this information was not transmitted to the vehicle behind. Another possibility is the tractor coming to a sudden stop due to some other problem, such as being rammed by a vehicle. Accidents due to mechanical defects in the machine are serious problems and can cause great damage. Otegui (2002), investigating an accident involving a tractor and a bus that ended in 15 fatalities, found that the accident was due to the failure of a trailer coupling, a result of the rapid deterioration of a screw that had been subjected to high loads causing a crack to form, something that could possibly be avoided with correct maintenance.

Groups 9 and 12, neighbours in the U-matrix projection, and very similar, are however different in relation to the time and road layout, Group 9 occurring on a curved road. Yotsutsuji *et al*. (2017) report that factors such as high speeds and sharp or long bends contribute to accidents on curved roads, for which the authors recommend a drastic reduction in speed when on a curve.

Group 10 was the only group in the entire study to include collision with a fixed object. This type of collision occurs when the tractor collides with objects that are not movable, such as trees and signs. The accidents in group 10 probably occurred due to the tractor presenting a problem, making it difficult to control and causing it to collide with a sign or other object on the shoulder.

It should be noted that once again the cause of the accidents was a mechanical defect in the vehicle. Special attention should be paid to this issue by offering operators in the region courses on the operation and maintenance of tractors. Alvarenga *et al*. (2017), conducting a study in the cerrado region of Minas Gerais, report that operators are rarely updated, resulting in an increase in accidents in the region. In addition, they recommend improving the quality of training courses so that operators can properly carry out their work.

The profile shown in Group 13 can be interpreted as occurrences where the type and cause were not identified by the Federal Highway Police, or even where the class was not addressed by the TAB, but was represented by a large number of incidents.

The profile of Group 15 may have been greatly influenced by the road layout. The driver probably entered the curve at a certain speed without realising that there was a tractor ahead travelling at a lower speed than his own vehicle. On noticing the machine, and trying to avoid a collision, the driver probably attempted to overtake and ended up colliding with the side of the tractor.

In Group 18, the type of accident was ‘transversal and rear-end collisions’. Transversal collisions commonly occur at intersections, and more rarely on straight roads. In the south-east, they were only represented by two groups despite this type of accident being very frequent according to Macedo *et al*. (2015a), who found that transversal collisions were the third most frequent type of tractor accident on federal highways in the state of Minas Gerais.

CONCLUSIONS

1. The classes that activated the most neurons were rear-end collision, lack of attention and other causes, no casualties, straight road, good weather conditions and Minas Gerais;
2. Using this technique, it was possible to profile various incidents, opening the possibility of analysing data related to accidents;
3. Using agricultural machinery at times of low natural lighting, i.e. late afternoon, night and during the early hours, should be avoided. Such conditions increase the risk of rear-end collisions and the severity of these accidents can be greater, even causing the death of those involved.

ACKNOWLEDGEMENTS

The authors wish to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the 16 th Superintendency of the Federal Highway Police.

REFERENCES

ALVARENGA, C. B. *et al.* Ocorrência de acidentes com máquinas agrícolas na região do cerrado de Minas Gerais. **Brazilian Journal of Applied Technology for Agricultural Science**, v. 10, n. 3, p. 27-36, 2017.

ANTUNES, S. M.; CORDEIRO, C.; TEIXEIRA, H. M. Analysis of fatal acidentes with tractor in the centre of Portugal: ten years analysis. **Forensic Science International**, v. 287, n. 6, p. 74-80, 2018.

BELLOCHIO, S. D. C. *et al.* “Estado da arte” sobre as publicações científicas envolvendo acidentes com tratores em vias públicas no Brasil. **Engenharia na Agricultura**, v. 26, n. 6, p. 541-554, 2018.

COMBERTI, L.; DEMICHELA, M.; BALDISSONE, G. A combined approach for the analysis of large occupational accident databases to support accident-prevention decision making. **Safety Science**, v. 106, n. 3, p. 191-202, 2018.

FERNANDES, H. C. *et al.* Acidentes com tratores agrícolas: natureza, causas e consequências. **Engenharia na Agricultura**, v. 22, n. 4, p. 361-371, 2014.

HAYKIN, S. **Redes neurais**: princípios e práticas. Porto Alegre: Bookman, 2001. 900 p.

HU, R. *et al.* ELM-SOM+: a continuous mapping for visualization. **Neurocomputing**, v. 3, n. 50, p. 1-10, 2019.

ISMKHAN, H. I-k-means-+: an iterative clustering algorithm based on an enhanced version of the k-means. **Pattern Recognition**, v. 79, n. 2, p. 402-413, 2018.

JI-HONG, R.; JIANG-CHENG, C.; NAN, W. Visual analysis of SOM network in fault diagnosis. **Physics Procedia**, v. 22, n. 1, p. 333-338, 2011.

KOHONEN, T. Essentials of the self-organizing map. **Neural Networks**, v. 37, n. 1, p. 52-65, 2013.

MACEDO, D. X. S. *et al.* Caracterização dos acidentes com máquinas agrícolas em rodovias federais no estado do Rio Grande do sul. **Ciência Rural**, v. 45, n. 1, p. 43-46, 2015b.

MACEDO, D. X. S. *et al.* Characterization of acidentes involving tractors in Brazilian federal highways in the state of Minas Gerais. **African Journal of Agricultural**, v. 10, n. 31, p. 3049-3055, 2015a.

MACEDO, D. X. S. *et al*. Uso de técnicas de agrupamento e rede neural em sinistros com máquinas agrícolas nas rodovias federais paranaenses. **Energia na Agricultura**, v. 33, n. 1, p. 14-21, 2018.

OLAWOYIN, R. *et al.* Application of artificial neural network (ANN)-self-organizing map (SOM) for the categorization of water, soil and sediment quality in petrochemical regions. **Expert Systems with Applications**, v. 40, n. 9, p. 3634-3648, 2013.

OLSON, R. H.; SOCCOLICH, S. A.; HANOWSKI, R. J. The impact of driver distraction in tractor-trailers and motorcoch buses. **Accident Analysis and Prevention**, v. 126, n. 5, p. 10-16, 2019.

OTEGUI, J. L. Fatigue damage leads to a serious traffic accident. **Engineering Failure Analysis**, v. 9, n. 1, p. 109-122, 2002.

PALAMARA, F.; PIGLIONE, F.; PICCININI, N. Self-organizing map and clustering algorithms for the analysis of occupational accident databases. **Safety Science**, v. 49, n. 9, p. 1215-1230, 2011.

SCHLOSSER, J. F. *et al.* Caracterização dos acidentes com tratores agrícolas. **Ciência Rural**, v. 32, n. 6, p. 977-981, 2002.

WANG, D. *et al*. Road traffic accident severity analysis: a census-based study in China. **Journal of Safety Research**, v. 70, n. 9, p. 135-147, 2019.

YOTSUTSUJI, H.; *et al.* A car-accident rate index for curved roads: a speed choise-based approach. **Transportation Research Procedia**, v. 25, p. 2108-2118, 2017.

ZHANG, G.; YAU, K. K. W.; ZHANG, X. Analyzing fault and severity in pedestrian-motor vehicle acidentes in China. **Accident Analysis and Prevention**, v. 73, n. 12, p. 141-150, 2014.