EFFICIENCY OF DISTINCT DATA MINING ALGORITHMS FOR CLASSIFYING STRESS LEVEL IN PIGLETS FROM THEIR VOCALIZATION

ALEXANDRA F. DA S. CORDEIRO[1], IRENILZA DE A. NÄÄS[2], STANLEY R. DE M. OLIVEIRA[3], FABIO VIOLARO[4], ANDRÉIA C. M. DE ALMEIDA[5]

ABSTRACT: Among the challenges of pig farming in today's competitive market, there is factor of the product traceability that ensures, among many points, animal welfare. Vocalization is a valuable tool to identify situations of stress in pigs, and it can be used in welfare records for traceability. The objective of this work was to identify stress in piglets using vocalization, calling this stress on three levels: no stress, moderate stress, and acute stress. An experiment was conducted on a commercial farm in the municipality of Holambra, São Paulo State , where vocalizations of twenty piglets were recorded during the castration procedure, and separated into two groups: without anesthesia and local anesthesia with lidocaine base. For the recording of acoustic signals, a unidirectional microphone was connected to a digital recorder, in which signals were digitized at a frequency of 44,100 Hz. For evaluation of sound signals, Praat® software was used, and different data mining algorithms were applied using Weka® software. The selection of attributes improved model accuracy, and the best attribute selection was used by applying Wrapper method, while the best classification algorithms were the k-NN and Naive Bayes. According to the results, it was possible to classify the level of stress in pigs through their vocalization.

KEYWORDS: vocal expression, well-being, level of pain, pig.

EFICIÊNCIA DE ALGORITMOS DISTINTOS DE MINERAÇÃO DE DADOS PARA A CLASSIFICAÇÃO DE NÍVEL DE ESTRESSE EM LEITÕES PELA SUA VOCALIZAÇÃO

RESUMO: Entre os desafios da suinocultura no atual mercado competitivo, destaca-se a rastreabilidade do produto, que garante, entre muitos pontos, a questão do bem-estar animal. A vocalização é uma ferramenta útil para identificar situações de estresse em suínos e pode ser usada em registros de bem-estar, em processos de rastreabilidade. Este trabalho teve o objetivo de identificar estresse em leitões através da vocalização, classificando esse estresse em três níveis: sem estresse, estresse moderado e estresse agudo. Foi realizado um experimento em granja comercial da cidade de Holambra-SP, onde se gravou a vocalização de vinte leitões durante o procedimento de castração, separados em dois grupos: sem anestesia e com anestesia local à base de Lidocaína. Para a captura dos sinais acústicos, foi utilizado um microfone unidirecional conectado a um gravador digital, em que os sinais foram digitalizados a uma frequência de 44.100 Hz. Para análises dos sinais sonoros, foi usado o *software* Praat®, e diferentes algoritmos de mineração dos dados foram aplicados no *software* Weka®. A seleção de atributos melhorou a acurácia do modelo, sendo que o melhor método de seleção de atributos usado foi o Wrapper, enquanto os melhores algoritmos de classificação foram o k- NN e o Naive Bayes. De acordo com os resultados, foi possível classificar o nível de estresse em suínos através de sua vocalização.

PALAVRAS-CHAVE: expressão vocal, bem-estar, nível de dor, suínos.

INTRODUCTION

Pork is the most consumed type of meat in the world, but to attract today's increasingly demanding consumers, the pig industry seeks competitiveness of its products. Price, quality and biosafety comprise the current success tripod in food trading. The requirements for this market are becoming increasingly stringent and follow predominantly the standards set by international bodies. Among these requirements, there is the subject of animal welfare.

Surgical castration is mandatory for male pigs aimed for slaughtering, however because it affects animal welfare, researchers seek less painful alternatives to this procedure. According to WHITE et al. (1995), local anesthesia can reduce the stress of pain caused by this procedure. However, there is disagreement on this issue (LEIDG et al., 2009). FITZPATRICK et al. (2006), indicate that methods to objectively measure pain and welfare in farm animals need to be very well studied in order to avoid incorrect results.

Many authors studied vocalization as a tool to measure animal welfare (MARX et al., 2003; NÄÄS et al., 2008; LEIDIG, et al., 2009). According to KRANENDONK et al. (2006), piglets of sows treated with cortisol, a major stress-related hormone, vocalized more than piglets from sows not treated with this hormone, when placed in a different environment. SILVA et al. (2008) developed an algorithm to identify the cough incidence site in the farm, using the arrival time of sound picked up by microphones distributed in different bays, and managed to locate the place of origin of the cough.

Data mining is a process of acquiring new knowledge from a database and is used as a technique for pattern recognition (HAN & KAMBER, 2006). The systematic collection of pig vocalization at different times during the castration procedure generates a supervised database of animal welfare associated with parameters that characterize each vocalization. From this database, it is possible to extract a pattern of well being according to the vocalization, using data mining techniques.

Thus, the objective of this study was to identify the level of stress in piglets from their vocalization.

MATERIAL AND METHODS

Description of the experiment

Vocalization was recorded during surgical castration in two groups of 10 animals each, aged between three and four days old. In the first group, castration was performed without anesthesia, and in the second group, with local anesthesia. The technique chosen for castration was the scrotal method, in which the animal was first contained in a wooden box specific for castration with the dimensions 0.10 m wide, 0.10 m high and 0.30 m length. With the animal in the supine, asepsis was performed in the scrotum area with iodine, followed by a vertical incision in each testicle with a scalpel, being exteriorized and pulled up to expose the spermatic cord until its rupture. After removal of the testes, the wound was cleaned with iodine.

The experimental protocol was approved by the UNICAMP Ethics Committee under the number 2224-1/2011. For the group of animals that received anesthesia, 20% Lidocaine anesthetic was applied ten minutes before testes incision. Vocalizations were recorded in different stress situations (Table 1) during surgery.

TABLE 1. Description of the stress conditions that the animals were exposed.

Procedure	Description
Picking	Holding the animal by hand
First weight	Weighing the animal before castration
Holding	Holding the animal in castration box
Anesthesia	Applying of the anesthetic in the testes
Castration	Performing the surgery
Post-anesthesia castration	Performing the surgery the animals previously anesthetized
Second weight	Weighing the animal after castration
Post-anesthesia second weight	Weighing the animal treated with anesthesia, after castration
Release	Return of the animals in the stalls of origin

For the recording of acoustic signals, it was used a system to capture the signals composed of a Yoga EM 9600® directional microphone (Figure 1A) positioned approximately 15 cm from the mouth of the animals. This microphone was connected to a Marantz PMD 660 digital recorder (Figure 1B) for digitization of voice signals at a frequency of 44,100 Hz.

a   **b**

FIGURE 1. System for registering the sound composed by a Yoga EM 9600®directional microphone (a) and a Marantz® PMD 660 digital sound recorder (b).

Signal Analyses

The sound signals recorded were edited and analyzed using Praat® software. Firstly, the noises such as human speech or other sounds from the vocalization of pigs were excluded. Each signal was divided into three samples and for each, the maximum and minimum amplitude, the Pith frequency, and frequency of formants 1; 2; 3 and 4 were extracted. In addition to these features, the duration of the signal, and the average of three samples for each attribute were obtained from the amplitude range (which is the difference between the maximum and minimum amplitudes) (Table 2). Through analysis of variance of the means ($p < 0.05$) performed in Minitab® software, it was possible to perform discretization of the values for 29 attributes (Table 2) in three ranges: low, medium and high. The stress situations were also classified into three levels (no stress, moderate stress and acute stress).

From these results, data mining was performed on Weka® Software. Initially, the attributes were ranked using *infogain* methods, which categorizes the attributes of information gain; *gainration,* which is also based on information gain, but this amount of information is sensitive to the targeted range of attribute values, and (CFS-correlation based feature selection), which selects attributes highly correlated with the class, and less correlated among themselves. From the selected attributes, J48 algorithms (decision tree) were executed based on the entropy of the attributes; k-NN (k-nearest neighbors), which determines the importance of a variable from the Euclidean distance;

Naive Bayes (Bayesian classifier), multilayer perceptron (neural networks), which learns strong connections for some inputs and weak correlations for others, and SVM (Support Vector Machine).

TABLE 2. Acoustic attributes and their units.

Attribute Number	Attribute Name	Unity	Attribute Number	Attribute Name	Unity	Attribute Number	Attribute Name	Unity
1	Signal Energy	Pa2*s	11	Formant 4 of Sample 1	Hertz	21	Formant 2 of Sample 3	Hertz
2	Signal Duration	s	12	Intensity of Sample 2	dB	22	Formant 3 of Sample 3	Hertz
3	Maximum Amplitude	Pa	13	Pith Frequency of Sample 2	Hertz	23	Formant 4 of Sample 3	Hertz
4	Minimum Amplitude	Pa	14	Formant 1 of Sample 2	Hertz	24	Mean Intensity	dB
5	Amplitude Interval	Pa	15	Formant 2 of Sample 2	Hertz	25	Mean Pith Frequency	Hertz
6	Intensity Sample 1	dB	16	Formant 3 of Sample 2	Hertz	26	Mean Formant 1	Hertz
7	Pith frequency of Sample 1	Hertz	17	Formant 4 of Sample 2	Hertz	27	Mean Formant 2	Hertz
8	Formant 1 of Sample 1	Hertz	18	Intensity of Sample 3	dB	28	Mean Formant 3	Hertz
9	Formant 2 of Sample 1	Hertz	19	Pith Frequency of Sample 3	Hertz	29	Mean Formant 4	Hertz
10	Formant 3 of Sample 1	Hertz	20	Formant 1 of Sample 3	Hertz			

To compare the performance of algorithms, it was used the Roc (Receiver Operating Characteristic) curve, a graphical tool used for algorithm evaluation. The graph obtained describes the relationship between rates of true positives (TP rate) and false positives (FP rate) of the classifier. The classifiers inner to the "convex hull" have a lower precision rate than those that are contained in the "convex hull". The greater the area under the curve, the higher the precision of the model.

RESULTS AND DISCUSSION

Analysis of variance was performed for each of the attributes of the acoustic signals evaluated, and showed difference ($p < 0.05$) in vocalizationof animals exposed to different conditions of stress (Figure 2) for the attribute representing variation of maximum and minimum signal amplitude. According to these results, there was no difference in the vocalization of piglets castrated with or without anesthesia ($p < 0.05$).

WHITE et al. (1995) found that piglets castrated without local anesthesia had a higher heart rate and vocalization with higher energy, suggesting that anesthesia could reduce castration stress. However, according to LEIDIG et al. (2009), local anesthesia may improve the well being of piglets during castration, but due to the period of animal manipulation, added to the stress of anesthesia application, this effect is uncertain. According to PUPPE et al. (2005), holding and castration altered the vocalization quality of the piglets, indicating detriment of welfare of these animals.

In this experiment, it was observed that the animals were relaxed, during weighing and release; on the other hand, they seemed highly excited at the time of holding, anesthesia and castration, and intermediately excited during the second weighing post-anesthesia and handling, which is in agreement with the vocalization results (Figure 2), suggesting three stress levels (Table 3).
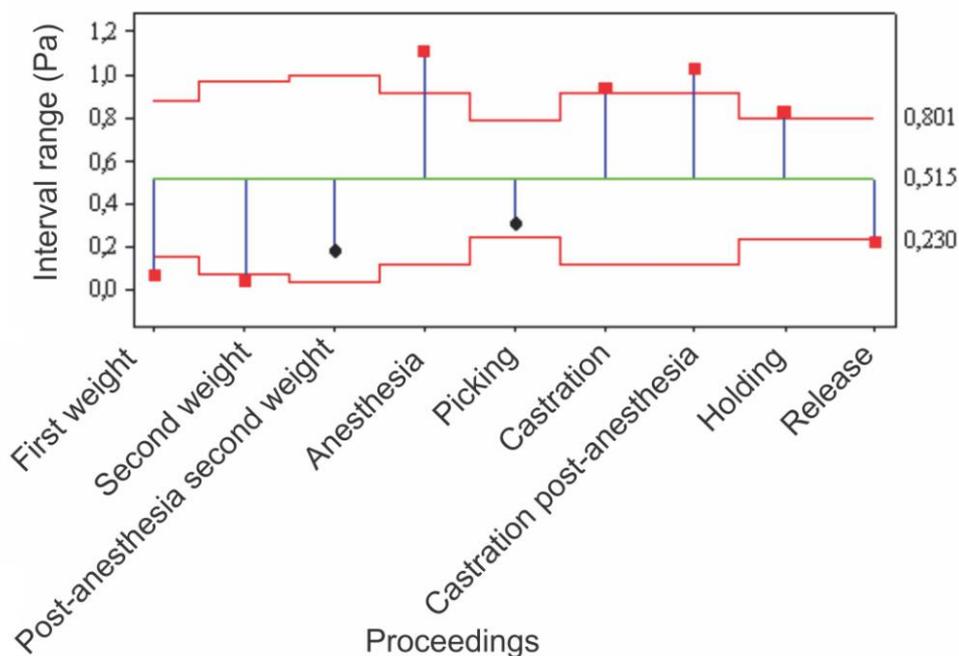
FIGURE 2. Schematic representation of the analysis of variance for the mean of the amplitude interval. The green line represents the data general mean, and the red lines indicate the inferior and superior limits. The black and red squares represent the mean values of management, and those in red color are outside of the limits and different of the general mean value; while those in black color are located within the limits and do not differ from the average values.

TABLE 3. Association between the stress condition and the corresponding stress level.

Stress Situation	Stress Level
1st weighing, 2nd weighing and release	No stress
Post-anesthesia 2nd weighing	Moderate stress
Anesthesia, castration, castration without anesthesia, and holding	Acute stress

From the ranking of attributes obtained by applying the Info Gain, Gain Ration, CFS and Wrapper methods (Table 4),Wrapper was the one that selected a fewer number of attributes, contributing to lower analysis costs.

TABLE 4. Selection and classification of the vocal signals attributes (29) for the distinct methods of the attributes selection.

Selection Approach	Selected Attributes
Info Gain	24; 2; 5; 4; 6; 3; 12; 18 and 28
Gain Ration	2; 5; 4; 24; 18; 3; 6 and 12
CFS	2; 5; 24 and 28
Wrapper- J48	5 and 28
Wrapper- k-NN	5; 10; 14; 17; 18; 20; 26 and 28
Wrapper- Naive Bayes	4; 5 and 28
Wrapper- Neural Networks	2; 5; 21; 25; 26 and 28
Wrapper- SMO	5; 26 and 28

Note: Ranking from left to right

Overall, there were differences in the selection of attributes among the distinct methods evaluated. However, the attribute Amplitude Interval presented the best ranking, indicating that it is the most important factor for predicting the stress level in piglets. The mean of the formant 3 and the signal duration also were ranked as important. MARX et al. (2003) evaluated the vocalization of pigs during castration and found that the vocalizations associated with pain can be identified and characterized mainly by the signal energy, frequency, and duration of the call. CORDEIRO et al. (2009) suggest that the energy of the signal increases as a function of the stress to which the animal is subjected. However, in the present study, only the signal energy was not a good parameter to predict the level of stress.

In general, the algorithm performance was best with the attributes selected by the Wrapper method than the ones selected by the other methods (Figure 3), thus being the best selection method. The highest accuracy (76.78%) was obtained by the k-NN and the attributes selected by the Wrapper method. To build this model, the algorithm used only eight of the 29 attributes.
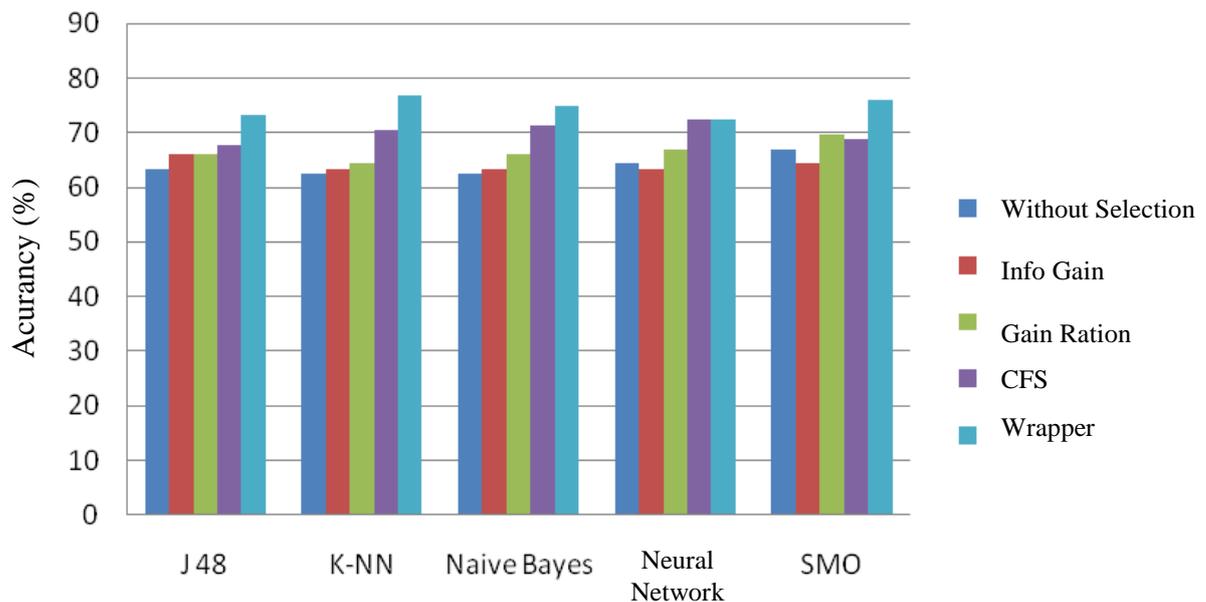
Algorithms and critheria for selection of attributes

FIGURE 3. Accuracy (%) found by the use of the distinct evaluated algorithms.

The accuracy for each stress level evaluated (Figure 4) indicates that, although the detection of three levels of stress are important, the cost is higher when the class without stress is mistaken, because a suffering animal cannot be classified as non-stressed. Thus, there is a need to achieve greater accuracy for this class. Observe that the no-stress class had greater accuracy (0.70) in the Naive Bayes algorithm, with the attributes selected by the Wrapper method. This model used only three attributes, with an overall accuracy of 75%.
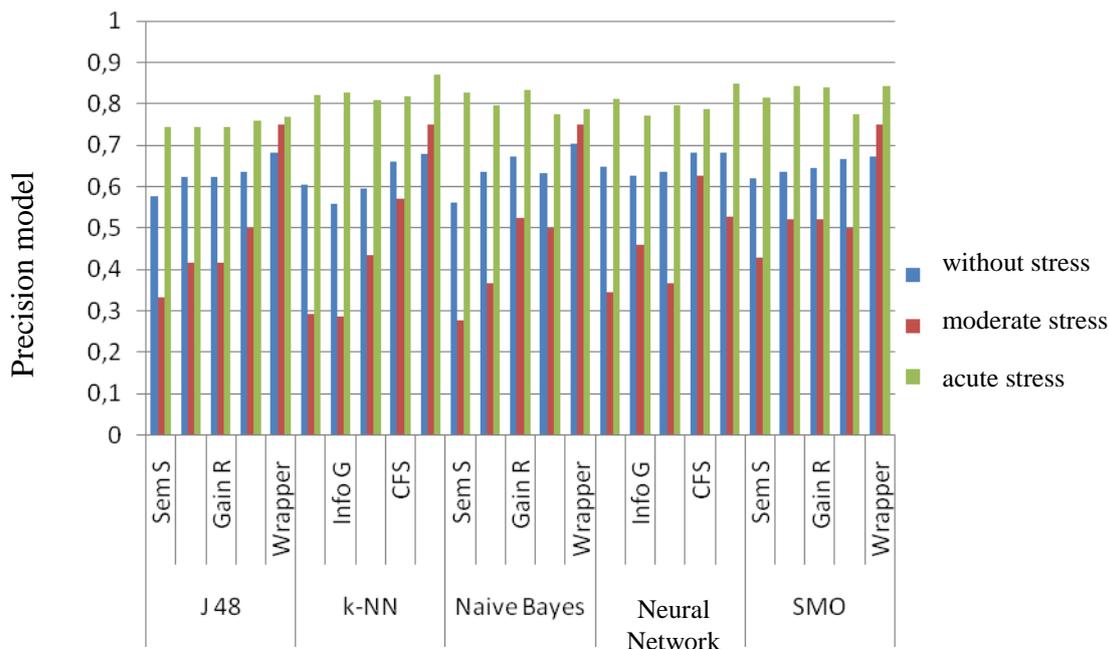
FIGURE 4. Precision level for each stage of stress of the distinct algorithms and selection methods.

The best classifiers were developed by k-NN and Naive Bayes (Figure 5), and the remaining ones were discarded, with k-NN being more efficient than Naive Bayes. However, it is noteworthy that the latter showed greater accuracy for the class without stress (0.705). Furthermore, implementation of this model has a lower computational cost, since it uses only three attributes: amplitude interval, mean of formant 3 and minimum amplitude.

In this research, the neural network was one of the algorithms that were discarded; however NÄÄS et al. (2008) developed an algorithm using artificial neural networks to classify vocalizations related stress.
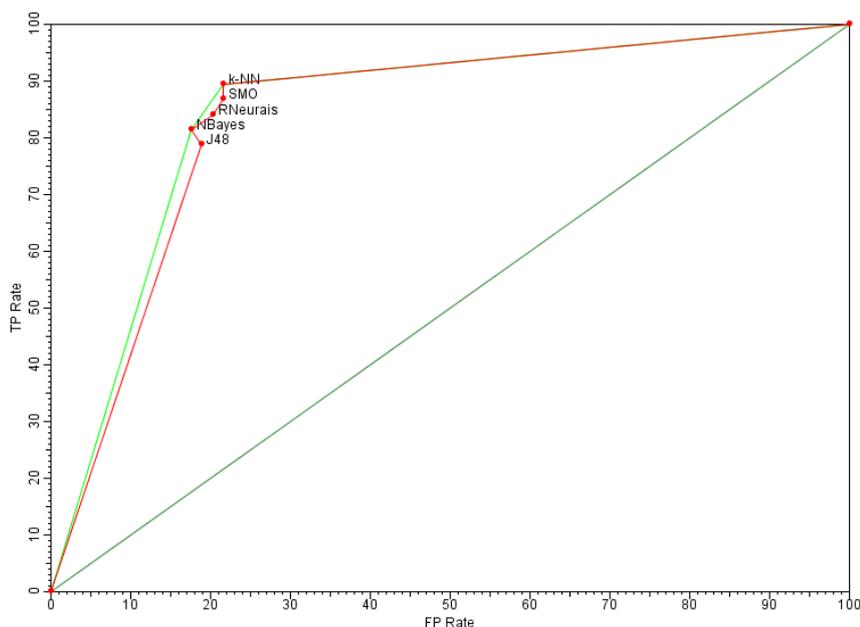
FIGURE 5. ROC curve considering the class without stress as the most important one.

Although the J48 algorithm was not be the best model developed, it achieved an overall accuracy of 75% and an accuracy of 0.682% for the class without stress, using only two attributes

and allowing illustration of the knowledge acquired, with a decision tree generated (Figure 6). According to the decision tree, the gap or amplitude difference is the most important attribute to predict the level of stress. If the amplitude range is high or medium, the stress is severe. But if the amplitude range is low, it will depend on the average of the formant 3, which is the average when the stress is moderate, and low or high when the animal is not under stress. RISI (2010) identified the intensity level of injury of arthritis in pigs through the vocalization of these animals by analyzing sound intensity, fundamental frequency and formants 1 and 2. According to the author, the mean values of these attributes tend to decrease as the level of injury increases. On the other hand, DUPJAN et al. (2008) found differences in resonance frequencies in pig vocalization due to aversive stimulus.
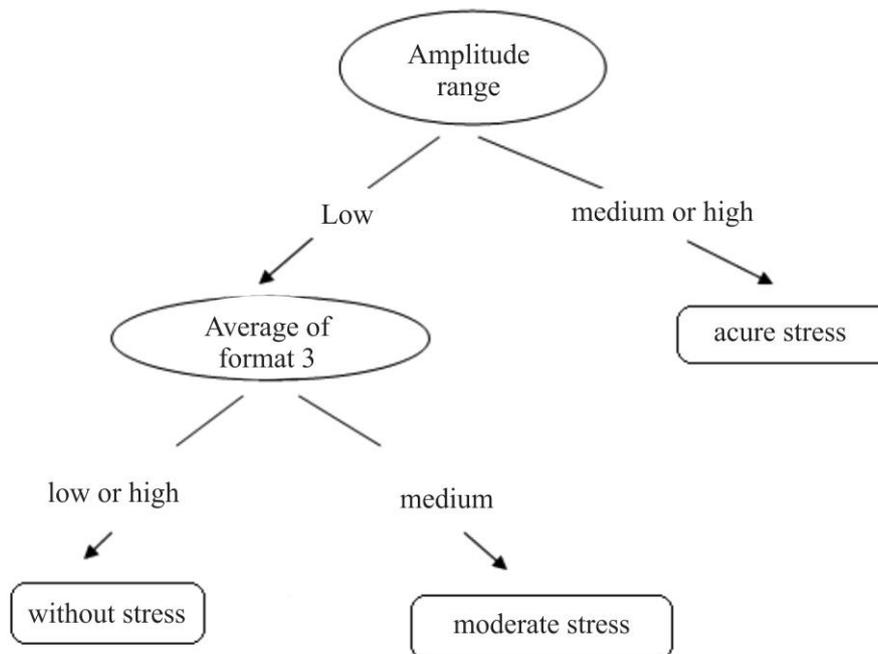
FIGURE 6. Decision tree obtained by the use of the Weka® software.

CONCLUSIONS

In this research, it was possible to classify the level of stress in pigs by their vocalization. The use of selection attributes enabled to reduce the number of sound characteristics to build the model, improving its accuracy.

ACKNOWLEDGMENTS

REFERENCES

CORDEIRO, A.F.S.; PEREIRA, E.M.; NÄÄS, I.A.; SILVA, W. T.; MOURA, D.J. Medida de vocalização de suínos (*Sus scrofa*) como um indicador de gasto energético. *Revista Brasileira de Engenharia de Biosistemas*, Campinas, v.2, n.2, p.143-152, 2009.

DUPJAN, S.; SCHÖN, P-C.; PUPPE, B.; TUCHSCHERER, A.; MANTEUFFEL, G. Differential vocal responses to physical and mental stressors in domestic pigs (*Sus scrofa*). *Applied Animal Behaviour Science*, Londres, v.114, n.1, p.105-115, 2008.

FITZPATRICK, J.; SCOTT, M.; NOLAN, A. Assessment of pain and welfare in sheep. *Small Ruminants Research*, Amsterdam, v.62, n.1, p.55-61, 2006.

HAN, J.; KAMBER, M. Data mining - concepts and techniques. 2nd ed. Nova York: Morgan Kaufmann, 2006.

KRANENDONK, G.; HOPSTER, H.; FILLERUP, M.; EKKEL, E.D.; MULDER, E.D. Cortisol administration to pregnant sows affects novelty-induced locomotion, aggressive behaviour, and blunts gender differences in their offspring. *Hormones and Behaviour*, Amsterdam, v.49, n.5, p.663 - 672, 2006.

LEIDIG, M.S.; HERTRAMPF, B.; FAILING, K.; SCHUMANN, A.; REINER, G. Pain and discomfort in male piglets during surgical castration with and without local anesthesia as determined by vocalization and defense behaviour. *Applied Animal Behaviour Science*, Amsterdam, v.116, n.2-4, p.174-178, 2009.

MARX, G.; HORN, T.; THIELEBEIN, J.; KNUBEL, B., BORELL, E. Analysis of pain-related vocalization in young pigs. *Journal of Sound and Vibration*, Amsterdam v.266, n.3, p.687-698, 2003.

NÄÄS, I.A.; CAMPOS, L.S.L.; BARACHO, M.S.; TOLON, Y.B. Uso de redes neurais artificiais na identificação de vocalização de suínos. *Engenharia Agrícola*, Jaboticabal, v.28, n.2, p.204-216, 2008.

PUPPE, B.; SCHÖN, P.C.; TUCHSCHERER, A.; MANTEUFFEL, G. Castration- induced vocalization in domestic piglets, *Sus scrofa*: Complex and specific alterations of the vocal quality. *Applied Animal Behaviour Science*, Amsterdam, v.95, n.1-2, p.67-78, 2005.

RISI, N. Uso da vocalização como indicador patológico em leitões na fase de maternidade. 2010. 93f. Dissertação - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, 2010.

SILVA, M.; FERRARI, S.; COSTA, A.; AERTS, J.M.; GUARINO, M.; BERCKMANS, D. Cough localization for the detection of respiratory diseases in pig houses. *Computers and Electronics in Agriculture*, Amsterdam, v.64, n.2, p.286-292, 2008.

WHITE, R.G.; DESHAZER, J.A.; TRESSLER, C.J.; BORCHER, G.M.; DAVEY, S., WANINGE, A.; PARKHURST, A.M.; MILANUK, M.J.; CLEMENS, E.T. Vocalization and physiological response of pigs during castration with of without a local anesthetic. *Journal of Animal Science*, Champaign, v.73, p.381-386, 1995.