CRISP-DM and K-means neutrosophic in the analysis of risk factors for hearing loss in children.

CRISP-DM y K-means neutrosofía en el análisis de factores de riesgo de pérdida de audición en niños.

Enriqueta B. Núñez Arias[1], Beatríz M. González Nuñez[2], Lisset Nonell Fernández[3], Jorge M. Rodríguez Pupo[4]

[1]Holguín Provincial University Hospital. Cuba. E-mail: queta@infomed.sld.cu
[2]Holguín Electric Company. Cuba. E-mail: beatriz.gonzalez3588@gmail.com
[3]Vladimir Ilyich Lenin University General Hospital. Holguín. Cuba.
[4]Lucía Íñiguez Landín Clinical Surgical Hospital. Holguin. Cuba. Email: jmichelrodriguez81@gmail.com

Resumen: Los trastornos auditivos son un problema de salud, asociado a trastornos emocionales y psicológicos secundarios, que dificultan la adaptación familiar y social. Se sugiere que la situación podría mejorar con un diagnóstico temprano. En la provincia de Holguín, Cuba, se realiza un exhaustivo estudio como parte de una extensa investigación cuyos resultados conforman un gran dato sobre los factores de riesgo causales. Se eligió la metodología CRISP-DM y K-means neutrosophic para la modelación de los datos y poder realizar una correcta segmentación de la información contenida sobre estos estudios realizados en una población infantil de 0 a 36 meses de edad del Hospital Universitario Provincial de Holguín. Ello permite la planificación y desarrollo de acciones. Se pretende que este estudio contribuya a la detección temprana del riesgo de padecer hipoacusia neurosensorial. Esto permitirá a los especialistas proponer estrategias en los diferentes niveles de gestión de la salud que promuevan la mitigación de este fenómeno para períodos posteriores.

Palabras clave: hipoacusia, factores causales, metodología CRISP-DM, K-means Neutrosófia.

Summary: Hearing disorders are a health problem, associated with secondary emotional and psychological disorders, which make it difficult for people to adapt to the family and society. It is suggested that the situation could be improved with an early diagnosis. In Holguín province, Cuba, an exhaustive study is carried out as part of an extensive research whose results make up a big data on the causal risk factors. The CRISP-DM and K-means neutrosophic methodology was chosen for the modeling of the data and to be able to carry out a correct segmentation of the information contained about these studies carried out on a child population from 0 to 36 months of age from the Holguín Provincial University Hospital. , which allows the planning and development of actions. It is intended that this study contributes to the early detection of the risk of suffering from sensorineural hearing loss. This will allow specialists to propose strategies at different levels of health management that promote the mitigation of this phenomenon for later periods.

Keywords: hearing loss, causal factors, CRISP-DM methodology, K-means neutrosophic.

1. Introduction

Currently, hearing loss is a health and social problem at the same time. Today there is a high prevalence in the child population, it is estimated that between 1-3 out of every 1000 newborns suffer a severe bilateral loss and 1 in every 100 have mild or moderate hearing disorders. Hearing loss at birth or during the first years of life (the critical or privileged period for language acquisition) affects language development, communication, and therefore intellectual development. Due to which there are multiple secondary emotional and psychological disorders, which make it difficult for them to adapt to family and society. It is suggested that the situation could be improved with an early diagnosis, which would make it possible to take advantage of the first years of life,[1, 2].
Based on this reality, the Joint Committee on Child Hearing and the National Institute for Health of Cuba, have recommended screening of the child population, as the only way for early detection and adequate treatment of children with hearing disorders, on everyone with sensorineural hearing loss (NSH) [1-3] and "child populations at risk" have been defined [4, 5] for the study of causal factors. Resulting that in Cuba there is a "high" incidence

of hearing disorders in a child population [6, 7]. For this reason, since 1984, a health program has been established for the early identification of children with these disorders where several medical specialties intervene. In the case of Holguín province, since 1985, there has been a provincial consultation in the Provincial University Pediatric Hospital, having a team of specialists in: neurophysiology, audiology, neurology, pediatrics, ophthalmology, neurodevelopment, genetics and speech therapy. This screening program was based on hearing risk factors.

This screening is based on tests carried out with expensive techniques such as the Auditory Evoked Potential of the Brainstem (PEATC) and the obtaining and interpretation of the records [6, 7]. In Cuba, the screening program was organized with a territorial approach, locating diagnostic technology and specialized personnel in a reference center, covering a wide sector of the child population. For example, in this center mentioned above, an average of 1000 to 1500 children can be treated during one year. At the same time, it can provide clinical care for a considerable number of cases, referred through other sources.[6-8].

The studies carried out on a population of children with hearing disorders from 0 to 36 months of age are handled in a big dataset containing information of the period between 2005-2019 and are constantly updated. Reason why its analysis is complex, in addition to the fact that it supports the variables of the study of an approximate population of 4,200 infants, investigated at the University Pediatric Hospital of Holguín, Cuba. It also describes the behavior of variables in the two stages of the program:

- Clinical pre-selection by multidisciplinary evaluation and,
- the realization of auditory brainstem potentials, stimulating monaural at 70 and 30 decibels respectively.

In this study, the electrophysiological hearing threshold is determined and hearing loss is classified. This information must be given an order that denotes planning, development and maintenance, so a data mining technique is relevant.

In order to process all the information in the Big Dataset and to discover knowledge in the data, the CRISP-DM methodology (Cross Industry Standard Process for Data Mining) will be applied, as it is the one used in [1, 2]. It is exposed in [2] that, its origins date back to 1999 when an important consortium of European companies proposed, based on different versions of KDD, the development of this new free distribution reference guide, divided into 4 levels of abstraction organized hierarchically into tasks ranging from the more general, to the most specific cases and organizes the development of a Data Mining project. It defines a life cycle focused on exploration and analysis where the succession of phases is not exactly rigid [9-11].

The authors of the research, based on their medical experience and trajectory as researchers, empirically defend the need to prioritize the causal factors of the hearing disorder and it is necessary to prove a priori that this order should not be absolutely conditioned by the frequency with which they affect it. This statement becomes more complex because psychosocial and biological phenomena intervene that give the process a certain subjectivity and make its analysis diffuse from the logic imposed by mathematical statistics. Therefore, it is necessary to apply an analysis using data mining methods including the fuzzy logic posed by neutrosophic science so that the idea that it is valid not to rank causal factors according to their frequency of incidence can be confirmed.

Neutrosophy is a new branch of philosophy which studies the origin, nature and scope of neutralities, as well as their interactions with different ideational spectra, created by Professor Florentin Smarandache[12]. With the use of classical statistics, the data formed by sharp numbers are known, in neutrosophic statistics the data has some indeterminacy, the data can be ambiguous, vague, imprecise, incomplete, even unknown. Instead of sharp numbers used in classical statistics, sets (which respectively approximate these sharp numbers) are used in neutrosophic statistics[13].

Because it is a phenomenon of a biological nature where each person behaves as a different entity, it is necessary to combine data mining techniques to extract the greatest amount of information. That is why the K-means technique is chosen in its neutrosophic version, which answers several questions, including uncertainty. With this technique, the detailed analysis of specific characteristics would be possible even when the criteria are quantifiable but the prediction of a behavior between groups of individuals is difficult.

So, it is established as a problem to be solved: how to carry out a correct segmentation of the information contained in the dataset of the studies carried out during the hearing screening of a child population between 0 and 36 months of age from the Holguín Provincial University Hospital, which allows planning and development of actions. Therefore, the objective of the article is to apply segmentation methods through data mining techniques that allow breaking down and using these results for predictive purposes. It is intended that this study contributes to the early detection of the risk of suffering from sensorineural hearing loss. This will allow specialists to propose strategies at different levels of health management that promote the mitigation of this phenomenon for later periods.

2 Materials and methods

The study will be structured according to the phases that expose [9-11] of the CRISP-DM methodology as explained below:

1. Understanding the "business": This initial phase focuses on understanding the objectives and demands of the project from a business perspective. It then turns that knowledge of the data into a definition of a data mining problem and a preliminary plan designed to achieve the objectives.
2. Understanding the data: Understanding the data takes care of the initial data collection and continues with the activities that allow you to first become familiar with the data, identify its quality problems, discover preliminary knowledge in the data, and / or discover interesting subsets to formulate hypotheses. In this phase, data sources that were not being used up to now (external sources) are also taken into account.
3. Data preparation: The data preparation phase covers all the activities necessary to build the final data set (the data that will be provided by the modeling tools). Preparation tasks include data selection, data cleaning, construction of new variables, data integration, and data formatting.
4. Modeling: During this phase, data mining techniques are applied to our data. Various modeling techniques are applied and the parameters of their use are fine-tuned to the optimum values. Some modeling techniques need specific requirements on the data format, which may lead us back to the data preparation phase.
5. Evaluation: In this case, the previous models are evaluated to determine if they are useful for business needs. At this stage the models are already built and should be of high quality from a data analysis perspective.
6. Deployment: The deployment phase involves exploiting the models within a production environment. The creation of a model is not generally the end of the project, since its creation is a living process within the decision-making process of an organization (it may be necessary to remake the model to take into account new knowledge in the future).
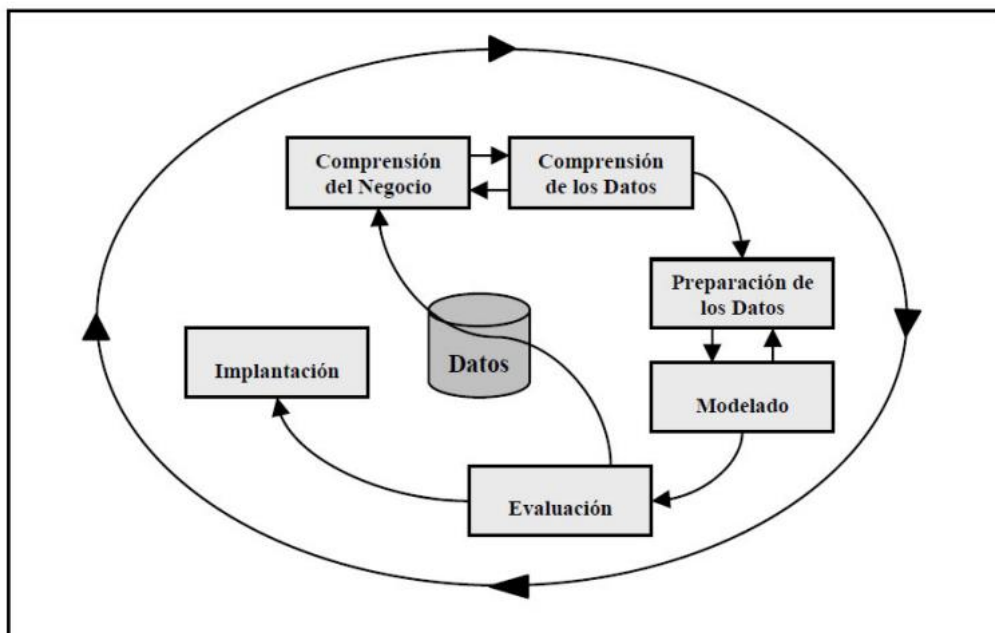

Figure 1. CRISP-DM process model [9]

The data modeling will be carried out with the Neutrosophic K-Means technique. The K-Means technique is used in Data Mining, due to its ease in handling and classifying large amounts of data through clustering. That is why the classical algorithm of the technique is appropriate due to the efficiency demonstrated for the decision-making process based on the interpretation of the linguistic terms provided by Neutrosophy. To better understand the technique, it will be explained below what both forms consist of:

K-Means: According [14-26] clustering means grouping things which are similar or have features in common, and so is the purpose of k-means clustering. K-means clustering is an unsupervised machine learning algorithm for clustering 'n' observations into 'k' clusters where k is predefined or user-defined constant. The main idea is to define k centroids, one for each cluster. The K-Means algorithm involves:

1. Choosing the number of clusters "k".
2. Randomly assign each point to a cluster.
3. Until clusters stop changing, repeat the following:
o   For each cluster, compute the cluster centroid by taking the mean vector of points in the cluster.
o   Assign each data point to the cluster for which the centroid is the closest.

Two things are very important in K-means, the first is to scale the variables before clustering the data, and second is to look at a scatter plot or a data table to estimate the number of cluster centers to set for the k parameter in the model.

Neutrosophy: it is a new branch of philosophy that studies the origin, nature and scope of neutralities created by Professor Florentin Smarandache. Its incorporation guarantees that the uncertainty of decision-making is taken into account, including indeterminacies where experts will issue their criteria evaluating linguistic and non-numerical terms, which constitutes the most natural form of measurement in human beings.[12, 13, 20, 27-29]. Logic and neutrosophic sets, for their part, constitute a generalization of Zadeh's logic and fuzzy sets, and especially of Atanassov's intuitionist logic, with multiple applications in the field of decision-making, image segmentation and machine learning.[12, 13, 30].

- o Definition 1[12, 31, 32]: Let X be a universe of discourse, a Neutrosophic Set (NS) is characterized by three membership functions, $u_A(x):, r_A(x):, v_A(x):: X \rightarrow]^-0, 1^+[$, which satisfy the condition $-0 \leq \inf u_A(x) + \inf r_A(x) + \inf v_A(x) \leq \sup u_A(x) + \sup r_A(x) + \sup v_A(x) \leq 3 + x \subset X \ \forall u_A(x), r_A(x)$ Y $v_A(x):$ denote the membership functions of true, indeterminate, and false of x in A, respectively, and their images are standard or non-standard subsets of $]^-0, 1^+[$.

Neutrosophic K-means: [20, 22, 33-35] is an extension of the classic K-Means, as a neutrosophic data mining technique for clustering. This analysis includes the diversity of the data and its fluctuation, since due to the proximity of the limits between them and their clusters of belonging, it is difficult to identify them, resulting in false conclusions and the existence of contradictions due to the uncertainty that this may generate. Based on what was stated by[36] The method consists in assigning to each data a value or degree of membership within each cluster (in this way the limits are smoothed and it is possible that a specific data may partially belong to more than one cluster)

- o Definition 2: Let X be the data set and xi an element, such that $X. x_i \in$
- o Definition 3: A partition $P = \{C_1, C_2, \ldots, C_c\}$ is said to be a soft partition of the data set X, if and only if it is true that: $(\forall x_i \in X, \forall C_j \in P) \leq \mu C_j(x_i) \leq 1$ and $(\forall x_i \in X, \exists C_j \in P)$ such that $\mu C_j(x_i) > 0$. Where $\mu C_c(x_i)$ denotes the degree to which xi belongs to the cluster Cj
- o Definition 4: It is said of a special soft partition when the sum of the degrees of membership of a specific point in all clusters is equal to 1 as shown in equation 1.

$$\sum_j \mu C_j(x_i) = 1 \ , (\forall x_i \in X \quad\quad\quad\quad o \quad (1)$$

- o Definition 5: A constrained soft partition is a partition that meets this additional condition. The Neutrosophic K-Means algorithm produces a constrained smooth partition and to do this the objective function J is extended in two ways:

  $\forall x_i \in X, \exists C_j \in P$ such that $\mu C_j(x_i) > 0$ where the degrees of neutrosophic membership of each data in each cluster are incorporated or;

  introducing an additional parameter that serves as exponent weight in the membership function, thus the extended objective function Jm is as shown in 2.

$$\mu c_1(x_1) = \frac{1}{\sum_{j=1}^{2} \left[\frac{\|x_1 - v_1\|^2}{\|x_1 - v_j\|^2} \right]^2} \tag{2}$$

Where P is a fuzzy partition of the data set X formed by $\{C_1, C_2, \ldots, C_k\}$ and the parameter m is a weight that determines the degree to which the partial members of a cluster affect the result.

Which refers to a similarity between the classical method and its neutrosophic extension, since the latter also tries to find a good partition by searching for the prototypes vi in such a way that they minimize the objective function Jm and that in the same way it must also look for the functions of membership that minimize Jm. μC_1

In addition to the method, equation 3 is established to calculate the initial membership functions of both clusters:

$$J_m(P, V) = \sum_{j=1}^{k} \sum_{j_k \in X} (\mu c_j(x_k))^m \|x_k - v_j\|^2 \tag{3}$$

The calculations are subsequently updated according to equation 4.

$$v_1 = \frac{\sum_{k=1}^{n} (\mu C_1(x_k))^2 x_k}{\sum_{k=1}^{n} (\mu C_1(x_k))^2} \tag{4}$$

In summary, the CRISP-DM methodology was taken as the common thread for this research, using the neutrosophic K-means for data modeling. Descriptive statistics will also be used in children with sensorineural hearing loss, who attended the provincial Audiology consultation at the "Octavio de la Concepción y de la Pedraja" Teaching Hospital, Holguín province, Cuba, during the years 2005-2019. As sources of information, the clinical records of the Audiology consultation file, the PEATC records of the neurophysiology consultation of patients in the previously specified age ranges were used.

The ages in months are chosen to better segment and describe the behavior of sensorineural hearing loss. We

worked with a chosen significance level of 5%. For the work with K-means, the Orange V 3.27.1 software is chosen. In the particular case of the neutrosophic K-means through a Python Script and the results are plotted in a Scatter Plot, as well as a Silohuette Plot to see in detail the Silohuette values in each cluster. And at the end of the process, a Data Table was placed to better analyze the results. To define the number of clusters (k), the method offered by the Orange k-Means widget was used, which allows executing several iterations and thus finding the best partition. The optimal value is the one with the highest Silohuette score, in this case 0.768, which corresponds to the option of k = 2 clusters.

To include the Neutrosophic part, it was necessary to program a Python script in which we applied formula 3 to calculate the initial membership functions of both clusters and formula 4 to adjust the calculations, iterating the process until the extended objective function is minimized, as expressed in equation 2. The Silohuette value was calculated from the Euclidean distance.

3 Results

Phase 1 and 2. Comprehension
It was taken as evidence of understanding of the study phenomenon (business) what was mentioned in section 1 of the article, as well as the following details:
- o Establishment of the bases for obtaining neurophysiological studies, the evoked potentials: collect signals on the scalp, calculate the voltage difference between the recording points, filter the signals in the frequency domain, amplify the signals, average segments, transduction of the collected signals, inscribing the signals.
- o Parameters to be evaluated: in neurophysiological studies (absolute latency of wave V and, replicability, morphology, amplitude, electrophysiological threshold)
- o Hearing Anatomy
- o PEATC generators and their general principles and interpretation
- o Application and advantages of the studies

Another analysis carried out to understand the data contained in big data is the fact that 63.5% of infants are male, therefore the data shown in the modeling may be influenced by gender. A review of the literature shows that boys take longer to develop language than girls, so it may be a variable to rule out in this analysis.
It is necessary to clean the data taking into account the phenomenon under study to discern among them which are relevant for analysis of the causal factors of risk of hearing loss in children aged 3-36 months. The partial conclusions reached are set out below:
- o Out of the total, 36.5% are female and 63.5% male, so the number of children who are candidates for audiological evaluation for presenting language delay is higher than in girls, which is in accordance with the literature consulted. According to [7, 8, 11, 37], in most patients gender variations are found because language development in boys is a slower process than in girls.
- o 36.5% of these were related to some genetic factor.
- o The age group that reported the most cases was found to represent less than 50% of the total number of cases studied in the 0 to 2-year-old group. Therefore, it is presumed that this low figure is due to the late diagnosis of hearing loss or those caused by postnatal factors such as: meningoencephalitis and progressive hearing loss.
- o The most frequent risk factors behaved in a similar way to the world reports: hypoxia with 20.6%, and family history of hearing loss of unknown cause and arterial hypertension during pregnancy, 19%. Perinatal conditions are associated with the presence of hearing loss, as they constitute prenatal risk factors such as: prematurity, toxic or infectious; arterial hypertension in pregnancy, hereditary causes, or those that occur from the moment of conception (peri- and postnatal risk factors).
- o According to what was stated by [38-40]the data coincide with the statement that audiological risk factor is difficult to define. It is the main cause of hearing loss in the population at risk. Due to the association of multiple factors to find the relationship between an isolated factor and the amount of hearing loss, it is difficult.

The previous data show a variability in the causes of sensorineural hearing loss according to the age groups studied, so it is considered appropriate to determine, as part of the analysis of the causal risk factors for hearing loss in children aged 3-36 months. What will result in a proactive action for the implementation of prophylactic programs.

Phase 3-5. Preparation, Modeling and Evaluation
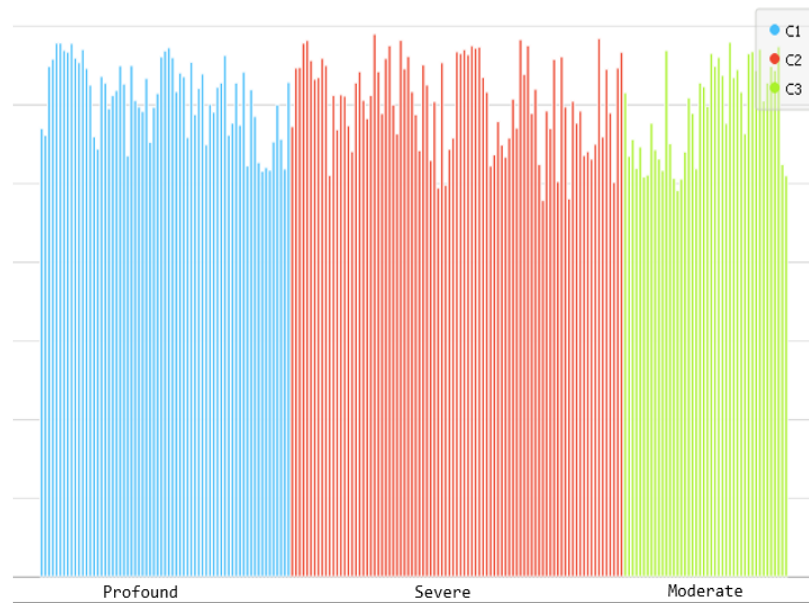In this phase, K-means is applied as a data mining technique. The following figures show the modeling of the data.

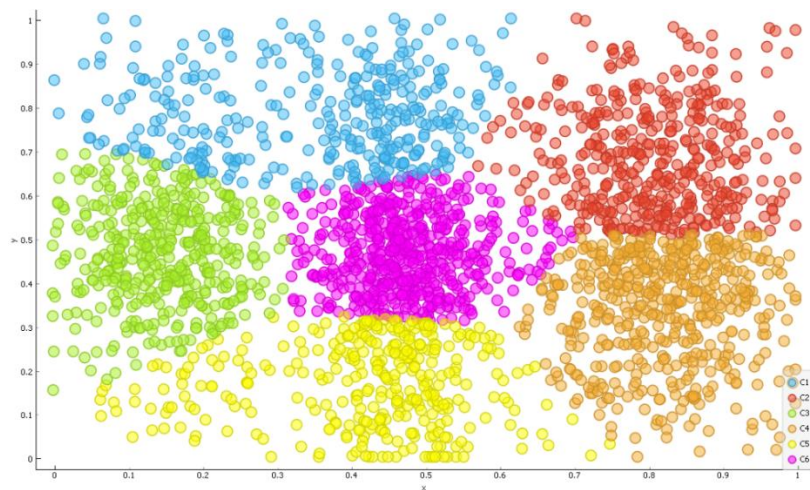**Figure 2.** Incidence by type of hearing loss

Figure 3. Clustering according to risk factors and their frequency according to Euclidean distance
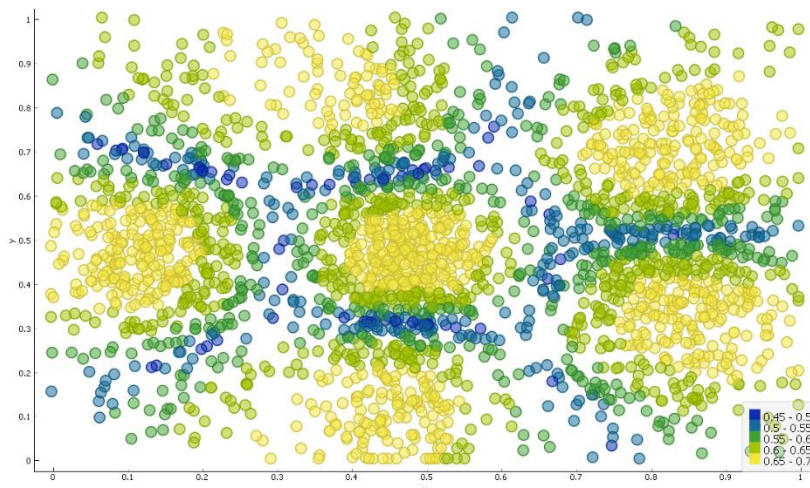

Figure 4. Fuzzy boundaries between clusters according to Euclidean distance.

As can be seen in figure 4, the limits between the sets are fuzzy, thus demonstrating the need for the application of Neutrosophy. This is useful to discern the level of membership of a data to a certain set and thus achieve correct conclusions according to the variability of the phenomenon. It is found that the audiological risk factor is difficult to define. It is the main cause of hearing loss in the population at risk. Due to the association of multiple factors to find the relationship between an isolated factor and the amount of hearing loss, it is difficult.

However, based on the calculation obtained with the previous graphs, it was possible to carry out a cross-sectional statistical study and recalculate the results, obtaining the following data on the incidence of the most frequent risk factors in the study. It was observed that there are clusters where a high variability between risk factors is demonstrated, but a pattern of behavior is shown according to its influence on the type of hearing loss detected. In the case of cluster 5, it can be seen that postnatal factors such as meningoencephalitis and progressive hearing loss extend throughout moderate and severe hearing loss.

When analyzing the results obtained, it can be verified that sensorineural hearing loss behaved in a similar way to world reports.
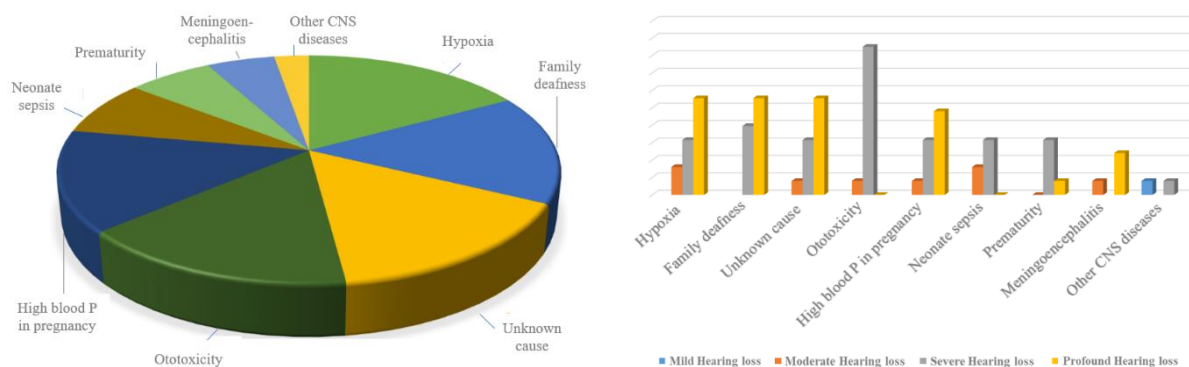

Figure 5. Distribution of the sample according to the most frequent risk factors.

Phase 6. Deployment
This phase involves the exploitation of the models within a production environment, that is, the strategies adopted as part of a higher study will be evaluated to determine their effectiveness. At the moment it will not be part of this investigation until this moment.

Conclusions

Hearing loss in children is nowadays a challenge for the health system. Despite the achievements on this field, there is still a need for earlier detection of this disability. This study shows that universal evaluation is the only truly effective alternative for the screening of congenital hearing loss. Educating the population about the audiological risk factors and the deeper training of health personnel, fundamentally, Primary Health Care workers, are decisive. Educating both parents and relatives in prenatal consultations would be a life strategy. It is considered that there is a probability of late diagnosis which causes the greatest hearing losses in children older than 2 years due to postnatal factors such as infections of the central nervous system.

It is proven that neonatal hearing loss screening programs require appropriate technology for diagnosis, and the possibility of early and effective intervention (prosthesis and cochlear implant). On the other hand, evidence is obtained early on the child's poor hearing. The limits between the sets are fuzzy, which is why the need for the application of Neutrosophy is demonstrated. This is useful to discern the level of membership of a data to a certain set and thus achieve correct conclusions according to the variability of the phenomenon. It is found that the audiological risk factor is difficult to define. It is the main cause of hearing loss in the population at risk.

A predominance of profound and severe hearing loss and to a lesser degree of the moderate type is observed, where the main causal factors are postnatal diseases. Perinatal conditions are associated with the presence of hearing loss, there are prenatal risk factors such as: prematurity, toxic or infectious; arterial hypertension during pregnancy, as hereditary causes, or those that occur from the moment of conception.

References

1. Y. Pérez Villa, M. Mena Ramírez, L. Chamlati Aguirre, I. Mora Maganab, and I. Gutiérrez Farfána, "Utilidad diagnóstica del reflejo acústico para predecir audición en población pediátrica. ," Acta Otorrinolaringol Esp, vol. 65(6), pp. 332-338, 2014.

2. A. Alvo, C. Der M, and P. Délano, "Tamizaje universal de hipoacusia en el recién nacido. ," HospClín Universidad de Chile, vol. 21(2), pp. 170-176, 2010.

3. S. Borkoski Barreiro, J. Falcón González, J. Bueno Yanes, J. Pérez Bermúdez, Z. López Cano, and A. Ramos Macías, "Resultados de un programa de detección precoz de la hipoacusia neonatal.," Acta Otorrinolaringología Española, vol. 62(2), pp. 92-96, 2013.

4. J. Marco and S. Mateu, "Detección precoz de la hipoacusia en recién nacidos.," Libro blanco sobre hipoacusia. Madrid: Ministerio de Sanidad y Consumo, 2013.

5. F. Núñez Batalla, G. Trinidad Ramos, J. Sequí Canet, V. Alzina De Aguilar, and C. Jáudenes-Casaubón, "Indicadores de riesgo de hipoacusia neurosensorial infantil.," Acta Otorrinolaringol Española, vol. 63(5), pp. 382-390, 2012.

6. E. Mijares Nodarse, D. Herrera Alonso, J. Gaya Vázquez, E. Santos Febles, M. Pérez Abalo, L. Méndez Alarcón, et al., "Cribado auditivo neonatal on potenciales evocados auditivos de estado estable a múltiples frecuencias. ," Acta Otorrinolaringol Española., vol. 62(2), pp. 87-94, 2011.

7. J. Delgado Domínguez and PAPPS Infancia y Adolescencia, "Detección precoz de la hipoacusia infantil," Rev Pediatra Aten Primaria, vol. 13(50), pp. 279-297, 2011.

8. R. Pérez Alcantud and V. Alcantud García, "Factores de riesgo de pérdida auditiva en lactantes 2009-2010," Multimedia, vol. 17(3), 2013.

9. J. A. Gallardo Arancibia. (2007, Metodología para el Desarrollo de Proyectos en Minería de Datos CRISP-DM. EPB 603 Sistemas del Conocimiento. Available: www.oldemarrodriguez.com

10. J. F. Vallalta Rueda, "CRISP-DM: una metodología para minería de datos en salud," Healthdataminer.com, 2020.

11. J. Villena Román, "CRISP-DM La metodología para poner orden en los proyectos," S|ngular, 2016.

12. [12]   M. Leyva Vázquez and F. Smarandache, "Neutrosofía: Nuevos avances en el tratamiento de la incertidumbre," Pons, Bruselas. , 2018.

13. F. Smarandache, "Introduction to Neutrosophic Statistics," Sitech & Education Publishing 2014.

14. J. Burkardt, "K-means clustering," Virginia Tech, Advanced Research Computing, Interdisciplinary Center for Applied Mathematics, 2009.

15. E. Florido, F. Martínez-Álvarez, and J. Aznarte, Metodología basada en minería de datos para el descubrimiento de patrones precursores de terremotos de magnitud media y elevada, 2015.

16. M. A. I. Salame Ortiz, B. C. Pérez Mayorga, D. E. Huera Castro, and T. S. Viteri Paredes, "Las técnicas de reproducción asistida en el Ecuador ¿hecho jurídico o acto jurídico?," Uniandes EPISTEME. Revista digital de Ciencia, Tecnología e Innovación, vol. 5, 2018.

17. Z. E. Mamani Rodríguez, "Aplicación de la minería de datos distribuida usando algoritmo de clustering k-means para mejorar la calidad de servicios de las organizaciones modernas," Facultad de Ciencias Matemáticas, Unniversidad Nacional Mayor de San Marcos, 2015.

18. S. F. Chamba Jiménez, "Minería de Datos para segmentación de clientes en la empresa tecnológica Master PC," Unniversidad de Loja, 2015.

19. M. A. Naoui, B. Lejdel, and M. Ayad, "Using K-means algorithm for regression curve in big data system for business environment," Revista Cubana de Ciencias Informáticas, vol. 14, pp. 34-48, 2020.

20. Q. Li, Y. Ma, and S. Zhu, "Single valued Neutrosophic clustering algorithm Based on Tsallis Entropy Maximization," MDPI, 2018.

21. J. C. Córdova Galarza, "Aplicación de técnicas de minería de datos para predecir la deserción de los estudiantes que pertenecen al colegio fisco-misional "San Francisco" de la ciudad de Ibarra"," Facultad de Sistemas Mercantiles, Universidad Regional Autónoma de Los Andes "UNIANDES", 2014.

22. B. B. Yousif, M. M. Ata, N. Fawzy, and M. Obaya, "Toward an optimized neutrosophic K-means with genetic algorithm for automatic vehicle license plate recognition (ONKM-AVLPR)," IEEE Access, vol. 8, pp. 49285-49312, 2020.

23. D. T. Pham, S. S. Dimov, and C. D. Nguyen, "Selection of K in K-means clustering," Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, vol. 219, pp. 103-119, 2005.

24. I. Dabbura, "K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks," 2018.

25. A. S. Ashour, A. R. Hawas, Y. Guo, and M. A. Wahba, "A novel optimized neutrosophic k-means using genetic algorithm for skin lesion detection in dermoscopy images," Signal, Image and Video Processing, vol. 12, pp. 1311-1318, 2018.

26. M. Rizwan Khan. (2018). K Means Clustering Algorithm & its Application. Available: https://medium.com/datadriveninvestor

27. F. Smarandache, "A Unifying Field in Logics: Neutrosophic Logic. Neutrosophy, Neutrosophic Set, Neutrosophic Probability: Neutrsophic Logic. Neutrosophy, Neutrosophic Set, Neutrosophic Probability: Infinite Study.," 2005.

28. J. L. a. F. S. Salmerona, "Redesigning Decision Matrix Method with an indeterminacy-based inference process. Multispace and Multistructure. ," Neutrosophic Transdisciplinarity (100 Collected Papers of Sciences), vol. 4, p. 151, 2010.

29. M. Ali, Shabir, M., Smarandache, F., and Vladareanu, L. , "Neutrosophic LA-semigroup Rings," Neutrosophic Sets and Systems, vol. 7, pp. 81-88, 2015.

30. W. B. Vasantha, I. Kandasamy, and F. Smarandache, "Algebraic Structure of Neutrosophic Duplets in Neutrosophic Rings < Z U I>,< Q U I> and < R U I.> " Neutrosophic Sets and Systems, , vol. 23, pp. 85-95, 2018.

31. W. V. Kandasamy and F. Smarandache, "Fuzzy Neutrosophic Models for Social Scientists.," Education Publisher Inc., (2013)

32. W. B. V. a. F. S. Kandasamy, "Fuzzy cognitive maps and neutrosophic cognitive maps. ," American Research Press., 2003.

33. L. Morissette and S. Chartier, "The k-means clustering technique: General considerations and implementation in Mathematica. ," Tutorials in Quantitative Methods for Psychology., vol. 9, 2013.

34. T. Velmurugan and T. Santhanam, "Performance evaluation of k-means and fuzzy c-means clustering algorithms for statistical distributions of input data points," European Journal of Scientific Research, vol. 46, pp. 320-330, 2010.

35. M. N. Qureshi and M. V. Ahamad, "An improved method for image segmentation using K-means clustering with neutrosophic logic," Procedia computer science, vol. 132, pp. 534-540, 2018.

36. F. J. Cedeño Moran, M. P. Marcillo Sánchez, L. D. Roman Barrezueta, and C. N. Martillo Avilés. (2019) K- Means Neutrosófico para la segmentación de mercado. Neutrosophic Computing and Machine Learning.

37. S. Borkoski Barreiro, J. Falcón González, J. Limiñona Cañal, and A. Ramos Macías, "Evaluación del muy bajo peso (1500g) al nacer como indicador de riesgo pera la hipoacusia neurosensorial. ," Acta Otorrinolaringol Española., vol. 64(6), pp. 403-408, 2013.

38. [O. Cañete and M. Torrente, "Evaluación del programa de detección precoz de hipoacusia en recién nacidos prematuros extremos (RNPE), experiencia hospital Padre Hurtado," Rev Otorrinolaringológica Cir Cabeza Cuello, vol. 71(2), pp. 117-122, 2011.

39. R. Cabanillas Farpón and J. Cadiñanos Bañales, "Hipoacusias hereditarias. Asesoramiento genético.," Acta Otorrinolaringol Española, vol. 63(3), pp. 218-229, 2012.

40. A. Rojas Godoy, O. Gómez Gómez, and F. Rivas Muñoz, "Cumplimiento de la normatividad vigente para la detección temprana de hipoacusia neonatal.," Rev Salud Pública, vol. 16(3), pp. 462-472, 2018.

41. Gómez, G. Á., Moya, J. V., Ricardo, J. E., & Sánchez, C. B. V. (2020). Evaluating Strategies of Continuing Education for Academics Supported in the Pedagogical Model and Based on Plithogenic Sets (Vol. 37). Infinite Study.

42. Vázquez, M. L., Estupiñan, J., & Smarandache, F. (2020). Neutrosofía en Latinoamérica, avances y perspectivas. Revista Asociación Latinoamericana de Ciencias Neutrosóficas. ISSN 2574-1101, 14, 01-08.