INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

The Implementation of the K-Medoid Clustering for Grouping Hearing Loss Function on Excessive Smartphone Use

Eri Wahyudi [a], Dwiny Meidelfi [b,*], Nofrizal [c], Zulfan Saam [c], Juandi [c]

[a] Alifah Padang Institute of Health, North Padang, Padang, 25164, Indonesia
[b] Department of Information Technology, Politeknik Negeri Padang, Pauh, Padang, 25164, Indonesia
[c] Doctoral Program Environmental Science, University of Riau, Tampan, Pekanbaru, 28293, Indonesia
Corresponding author: *dwinymeidelfi@pnp.ac.id

Abstract— **During the current pandemic, smartphones have become a means of learning for all students in Indonesia, including high school students. Students use smartphones to send assignments, learn via video calls, and conduct online exams. The prolonged use of smartphones, from the beginning of learning hours in the morning to study hours in the evening, has a terrible impact on the ear health of high school students in Padang. Excessive smartphone use caused a decrease in the student's hearing function. Therefore, this study aims to group the audiometry results of high school students in Padang who have a hearing loss function. The audiogram result is only performed as the result of a frequency test of the subject's hearing in both the left and right ear. Conventionally, an otolaryngologist concluded the final decision of hearing loss ability. This research proposed an automatic classification of audiometry results using machine learning methods. The K-Medoids clustering was selected to classify the audiometry data in this research. Of 210 audiometry data, 91 data is confirmed by an otolaryngologist as valid data. By using the K-Medoids clustering, 93 data is classified into Normal hearing, Mild Hearing loss, and Moderate Hearing loss. The proposed model successfully grouped the audiometry data into three categories. The confusion matrix is applied to measure the model performance, which has 28,3% accuracy, 64,3% precision, and 21,4% recall.**

Keywords— **Students; ears; smartphones; clustering; K-Medoids; confusion matrix.**

I. INTRODUCTION

Currently, the development of smartphone technology is experiencing very rapid progress. As time passes, the number of smartphone users, ranging from children to parents, is increasing. A smartphone can support the performance of almost all human activities. Start sending messages, calls, chats, games, and more. The existence of this smartphone is considered so vital that it is not uncommon for most people to have a smartphone today.

In the digital era, smartphones have become the primary need of urban people to communicate, find information, and meet other needs. People can unknowingly be attached to their smartphones [1], [2], [3]. Concerning users' responses of focus and enjoyment, habit is one of the factors contributing to smartphone addiction. This smartphone addiction has a destructive impact on health [4], [5], one of which impacts the ears. Keeping ears clean and healthy is essential to prevent various ear problems, including ear infections, tinnitus,

deafness, and sudden deafness. Ear problems not only affect hearing, but they can also disrupt the balance of the body. Some variances were recognized in the audiograms of employees in the different professional categories, and in almost all categories, we discovered worse hearing in the left ear. Only a few publications in the medical literature have identified that the left ear has worse hearing levels and is considered significant information concerning the asymmetry of occupational noise-induced hearing loss [6]. Therefore, for the auditory and vestibular organs to continue to function correctly, the health of the ears must be appropriately maintained.

We are investigating the impact of smartphone use on high school students, but data processing is still manual. This could be a challenge if the stored data is quite significant in number, which requires a system to categorize the data for the students who have been impacted by smartphone use. When determining hearing loss, there was a discrepancy in the frequency and stage of threshold shift employed and the suggested course of action once hearing loss was identified

[7]. Figure 1 shows six categories of audiogram results [8]: Normal Hearing Ability, Mild Hearing Loss, Moderate Hearing Loss, Moderately Severe Hearing Loss, Severe Hearing Loss, and Profound Hearing Loss. During the Covid-19 pandemic, there are connections between daily verified Covid-19 cases and smartphone use. The compelling number is using smartphones' memory, Wi-Fi, and network switches. The most robust connection is between the smartphone use pattern and the COVID-19 case, subsequently disclosing the state of an outbreak [9], [10].
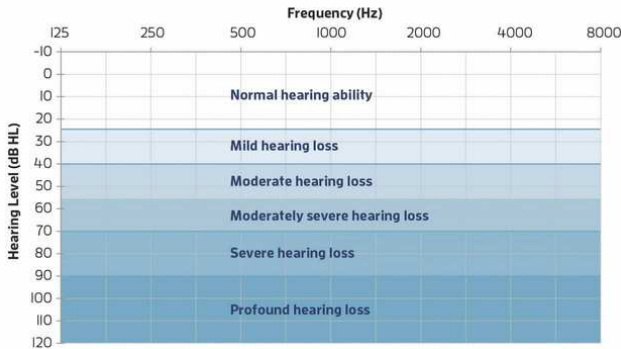

Fig. 1 Hearing loss Category

Recognizing the advantages of using hearing aids in quiet environments can be helpful for people with fewer communication or listening demands to decide on a hearing aid [11]. Many attempts have been made to advance unbalanced data classifiers due to their significance and data classification complexity [12]. Cluster analysis is the ideal method for database analysis since the average values of a database that contains variable and considerably varied data cannot be considered exemplary [13].

Clustering is a method or algorithm for grouping data or objects based on similarities [14]. This is data mining or part of data mining to obtain compelling patterns within voluminous data. This method can be used to classify data on high school students affected by smartphone use on ear health. Hence, in classifying cluster categories, the K-medoid clustering technique is considered [15], [16]. The presence of K-Medoids minimizes the overall distance between objects in each cluster. There are two stages in the algorithm's progression. Initially, it is obtained by finding k representative objects through an iterative selection of representative objects [17]. Both algorithm operation of K-Medoids clustering and K-Means clustering is initiated by randomly selecting k starting medoids representing the k clusters [18].

As the most significant unsupervised learning problem, clustering, like all other problems of this type, concerns identifying a structure in a set of unlabeled data. Clustering can be considered the most important unsupervised learning problem, so, as with every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. Broadly, clustering can also be defined as "The process of grouping objects into groups whose members are related in some way. "On the other hand, it can be characterized as a group of objects that are "similar" to one another and "dissimilar" to those found in other clusters. In addition, clustering contrasts with classification, where the objects are assigned to predetermined classes. The main advantage of clustering is that it allows for the direct discovery of compelling patterns and structures from massive data sets with little to no prior knowledge. Thus, the results are arbitrary and reliant on the implementation [19]. Clustering in text summarization aims to identify the significant subjects and subtopics in the document or collection of documents [20].

The reduced confusion matrix, a unique matrix, is the outcome of class grouping using the reduction method [21]. The number of instances assigned to each class can be seen in the confusion matrix, and the purpose is to calculate the classification accuracy [22]. To determine the final classification result, the outputs of various classifications are arranged in a decision-level fusion system by their confusion matrices [23]. Confusion matrix is considered an effective technique for assessing how well a classifier can distinguish between tuples of various classes, which is typically used to quantify the accuracy of a classifier or predictor [24].

II. Materials and Methods

A. Data Collections

The data collection was started by screening the student. The screening was done by asking the students to fill out the questionnaire. The questionnaire topics are about age, study habits, duration of smartphone usage, and visual ability. Based on those questionnaires, the medical staff decided on students who continue to test with audiograms. Mostly, the selected students must use a smartphone for more than two hours and have visual ability problems.

The audiometry test was located in "UPTD Keselamatan dan Kesehatan Kerja" Padang, which is a unit or division responsible for implementing and overseeing workplace safety and health measures within a specific region or jurisdiction. The test is supervised by an otolaryngologist who treats issues in the ears, nose, and throat (ENT Doctor). Based on the first screening, 210 recorded audiometry were achieved. Finally, 93 audiometry data is selected as the dataset for training data and testing the model. An otolaryngologist supervises the selection process.

B. K-Medoids

The K-Medoids Algorithm is one of the partitional clustering techniques to reduce the distance between a cluster's labeled points and its designated center. The K-Medoids algorithm differs significantly from the K-Means algorithm in several ways, as it selects data points as the center (medoids). Therefore, this study is carried out using the purity value in different data formats to evaluate the K-Medoid algorithm but with alternative data formats in order to determine better clustering outcomes (from several different data formats) [25].

The K-Medoids algorithm is employed to locate Medoids in a cluster at its central location [26]. As a classic partitioning technique or clustering method, K-Medoids divide object datasets into k groups based on prior knowledge. In a comparison of K-Medoids with K-Means, it was found that K-Medoids are stronger at handling noise (noise) as well as outliers (outliers) due to the reduction on some paired dissimilarities, instead of the sum of squares of Euclidean distances. A medoid is located in the cluster's center and can also be viewed as a cluster or as an entity in a cluster with the

smallest average difference between all objects. The flowchart of the K-Medoids algorithm can be seen in Figure 1.

Euclidean Distance size equation is applied to allocate each data (object) to the closest cluster[27] with the equation:

$$d_{euc}(x_{ij}, c_{kj}) = \sqrt{\sum_{j=1}^{p} \sum_{i=1}^{n} (x_{ij} - c_{kj})^2} \qquad (1)$$

Description:

$d_{euc}(x_{ij}, c_{kj})$ = Euclidean Distance between the i-th observation of the j-th variable to the center of the k-th cluster on the j-th variable

x_{ij} = object on the i-th observation on the j-th variable

c_{kj} = center of the k-th group on the j-th variable

p = the multiplicity of observed variables

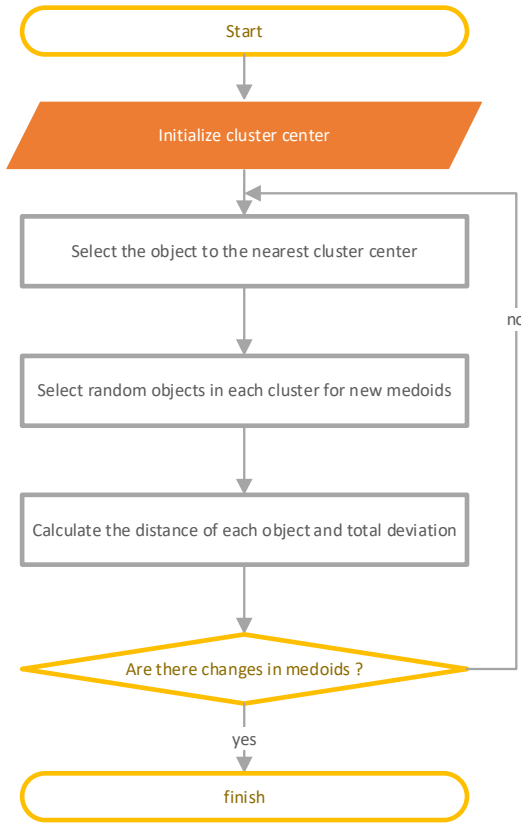n = the abundance of observations observed

Fig. 2 Flowchart of the K-Medoids algorithm

C. Confusion Matrix

We employ the Confusion Matrix method to evaluate the classification system's effectiveness in accounting for the accuracy of the results from the two devices [28]. In a multi-class classification task, when the confusion matrix is utilized as a tool to assess the performance of each instance labeled as one class by quantifying the classification overlap, it becomes prominent [29].

In the confusion matrix, the performance of machine learning classification models is commonly measured in tables and describes in detail whether the data set is correct or incorrect. This method is also one of the predictive analytics tools that displays and compares real or real values with the values of predictive models and generates matrices such as Accuracy, Precision, Recall, and F1-Score or F-Measure.

1) Accuracy: Accuracy is a test method determined by the proximity degree of the projected value and the actual value, and its prediction results can be measured when the correct classified data set is established. Accuracy involves the degree of similar results even after experiencing frequent tests and is not a category for boundary values [30].

2) Precision: Precision is a test method to determine the relevance of information received and obtained from the system by comparing the total amount of information from the two sources.

3) Recall: Recall is a test method carried out by comparing the total amount of applicable information received with the information in the information set (both for retained and discarded information).

4) F1-Score or F-Measure: The F1-Score value, code-named F-Measure, is determined by looking at the expected and actual result categories regarding precision and recall results.

TABLE I
CONFUSION MATRIX MEASUREMENTS

		Prediction Class	
		1	0
Actual Class	1	TP	FN
	0	FP	TN

Description:

TP (True Positive) = The precise amount of class 1 documents is designated as class 1

TN (True Negative) = The precise amount of class 0 documents is designated as class 0

FP (False Positive) = the number of class 0 documents that are misclassified as class 1

FN (False Negative) = the number of class 1 documents that are misclassified as class 0

The confusion matrix formula for calculating accuracy, precision, and recall is as follows.

$$accuracy = \frac{TP+TN}{Total} \qquad (2)$$

$$precission = \frac{TP}{TP+FP} \qquad (3)$$

$$recall = \frac{TP}{TP+FN} \qquad (4)$$

III. RESULT AND DISCUSSIONS

The sample in this research is part of the population determined from the best senior high schools and least favorite senior high school clusters. Based on the cluster, there were six senior high schools in Padang: SMA 1, SMA 3, SMA 5, SMA 6, SMA 8, and SMA 10. Samples were taken by using accidental sampling in each school, which became the research target of as many as 35 male and female students with a total sample of 210. The 210 total samples were re-netted to see the behavior and intensity of the smartphone used. For the results obtained in more than 4 hours,

measurement was carried out because it was in the heavy category with a total sample of 93 students.

A. Category Page

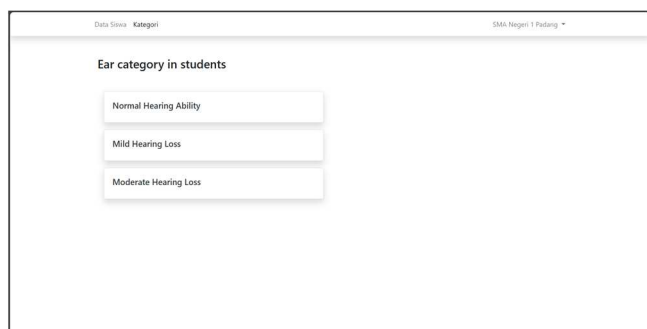Figure 3 shows the categories of damage to the student's ears.

Fig. 3 Category Page

B. Student Detail Page

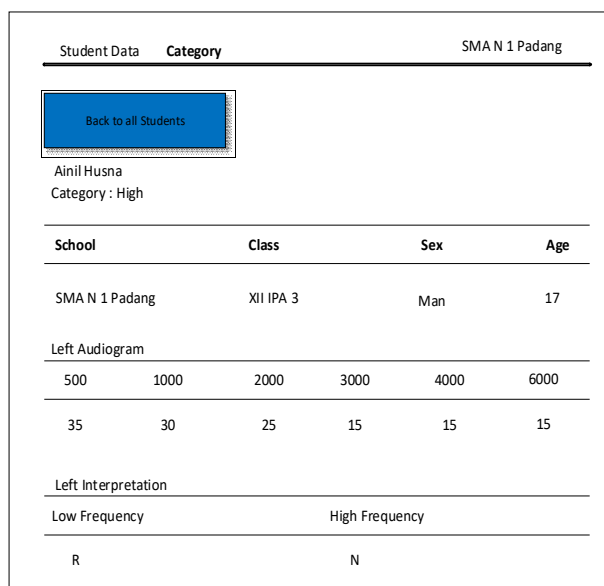Figure 2 shows the student's identity along with the results of the audiogram examination and interpretation.

Fig. 4 Student Detail Page

C. Case Studies

Tables 2 and 3 show audiogram data from SMA 1 Padang students.

TABLE II
STUDENT DATA OF SMA 1 PADANG

| Audiogram (db A) Left | | | | | | Left Interpretation | |
| Degree Evaluation (Hz) | | | Preventive Evaluation (Hz) | | | Low Frequency | High Frequency |
500	1000	2000	3000	4000	6000		
25	25	20	15	10	5	N	N
20	20	10	5	5	10	N	N
25	20	10	5	10	10	N	N
30	20	5	5	5	25	N	N
30	20	15	10	10	15	N	N
20	20	15	30	25	30	N	T
15	15	15	10	15	10	N	N

| Audiogram (db A) Left | | | | | | Left Interpretation | |
| Degree Evaluation (Hz) | | | Preventive Evaluation (Hz) | | | Low Frequency | High Frequency |
500	1000	2000	3000	4000	6000		
20	20	15	20	20	10	N	N
20	20	25	40	50	45	N	T
25	25	10	20	20	15	N	N
30	25	10	10	15	5	N	N
10	10	5	5	5	20	N	N
20	15	5	20	25	15	N	N
45	35	35	15	15	25	R	N
35	30	25	15	15	15	R	N

TABLE III
ADVANCED DATA OF SMA 1 PADANG STUDENTS

| Audiogram (db A) Right | | | | | | Right Interpretation | |
| Degree Evaluation (Hz) | | | Preventive Evaluation (Hz) | | | Low Frequency | High Frequency |
500	1000	2000	3000	4000	6000		
30	25	15	10	10	10	N	N
30	20	15	10	10	15	N	N
20	20	10	20	10	10	N	N
30	25	15	10	10	15	N	N
30	25	15	10	5	10	N	N
35	30	20	25	20	10	R	N
20	20	15	10	25	10	N	N
25	20	15	20	20	10	N	N
30	20	30	35	45	40	R	T
10	20	10	10	20	20	N	N
30	25	15	20	20	10	N	N
10	15	10	10	5	15	N	N
30	25	10	10	20	20	N	N
35	30	20	20	20	15	R	N
35	25	25	15	10	10	R	N

TABLE IV
TRANSFORMATION DATA

| Audiogram (db A) Left | | | | | | Left Interpretation | |
| Degree Evaluation (Hz) | | | Preventive Evaluation (Hz) | | | Low Frequency | High Frequency |
500	1000	2000	3000	4000	6000		
25	25	20	15	10	5	2	2
20	20	10	5	5	10	2	2
25	20	10	5	10	10	2	2
30	20	5	5	5	25	2	2
30	20	15	10	10	15	2	2
20	20	15	30	25	30	2	3
15	15	15	10	15	10	2	2
20	20	15	20	20	10	2	2
20	20	25	40	50	45	2	3
25	25	10	20	20	15	2	2
30	25	10	10	15	5	2	2
10	10	5	5	5	20	2	2
20	15	5	20	25	15	2	2
45	35	35	15	15	25	1	2
35	30	25	15	15	15	1	2
Min 10	10	5	5	5	5	1	2
Max 45	35	35	40	50	45	2	3

From the data shown above, it can be said that initially, the data must be transformed, specifically how the data must be converted into numbers as the data is processed using the K-

Medoids algorithm, which uses numerical data and looks for the highest and lowest values of the data. In the interpretation, the R-value (Normal Hearing Ability) becomes 1, the N-value (Mild Hearing Loss) becomes 2, and the T-value (Moderate et al.) becomes 3. It is shown in Tables 4 and 5.

TABLE V
ADVANCED DATA TRANSFORMATION

Audiogram (db A) Right						Right Interpretation	
Degree Evaluation (Hz)			Preventive Evaluation (Hz)			Low Frequency	High Frequency
500	1000	2000	3000	4000	6000		
30	25	15	10	10	10	2	2
30	20	15	10	10	15	2	2
20	20	10	20	10	10	2	2
30	25	15	10	10	15	2	2
30	25	15	10	5	10	2	2
35	30	20	25	20	10	1	2
20	20	15	10	25	10	2	2
25	20	15	20	20	10	2	2
30	20	30	35	45	40	1	3
10	20	10	10	20	20	2	2
30	25	15	20	20	10	2	2
10	15	10	10	5	15	2	2
30	25	10	10	20	20	2	2
35	30	20	20	20	15	1	2
35	25	25	15	10	10	1	2
Min	10	15	10	10	5	10	1	2
Max	35	30	30	35	45	40	2	3

From the data that has been transformed into numeric and look for the minimum and maximum, then normalize the data by:

$$normalize = \frac{(n - n_{min})}{(n_{max} - n_{min})} \qquad (5)$$

Description:
n = transformed data
n_{min} = minimal data from all data per column
n_{max} = max data from all data per column

Table 6 shows normalized transformation data, and table 7 shows advanced normalized transformation data.

TABLE VI
NORMALIZATION DATA

Audiogram (db A) Left						Left Interpretation	
Degree Evaluation (Hz)			Preventive Evaluation (Hz)			Low Frequency	High Frequency
500	1000	2000	3000	4000	6000		
0,429	0,600	0,500	0,286	0,111	0,000	1,000	0,000
0,286	0,400	0,167	0,000	0,000	0,125	1,000	0,000
0,429	0,400	0,167	0,000	0,111	0,125	1,000	0,000
0,571	0,400	0,000	0,000	0,000	0,500	1,000	0,000
0,571	0,400	0,333	0,143	0,111	0,250	1,000	0,000
0,286	0,400	0,333	0,714	0,444	0,625	1,000	1,000
0,143	0,200	0,333	0,143	0,222	0,125	1,000	0,000
0,286	0,400	0,333	0,429	0,333	0,125	1,000	0,000
0,286	0,400	0,667	1,000	1,000	1,000	1,000	1,000

Audiogram (db A) Left						Left Interpretation	
Degree Evaluation (Hz)			Preventive Evaluation (Hz)			Low Frequency	High Frequency
500	1000	2000	3000	4000	6000		
0,429	0,600	0,167	0,429	0,333	0,250	1,000	0,000
0,571	0,600	0,167	0,143	0,222	0,000	1,000	0,000
0,000	0,000	0,000	0,000	0,000	0,375	1,000	0,000
0,286	0,200	0,000	0,429	0,444	0,250	1,000	0,000
1,000	1,000	1,000	0,286	0,222	0,500	0,000	0,000
0,714	0,800	0,667	0,286	0,222	0,250	0,000	0,000

TABLE VII
ADVANCED DATA NORMALIZATION

Audiogram (db A) Right						Right Interpretation	
Degree Evaluation (Hz)			Preventive Evaluation (Hz)			Low Frequency	High Frequency
500	1000	2000	3000	4000	6000		
0,800	0,667	0,250	0,000	0,125	0,000	1,000	0,000
0,800	0,333	0,250	0,000	0,125	0,167	1,000	0,000
0,400	0,333	0,000	0,400	0,125	0,000	1,000	0,000
0,800	0,667	0,250	0,000	0,125	0,167	1,000	0,000
0,800	0,667	0,250	0,000	0,000	0,000	1,000	0,000
1,000	1,000	0,500	0,600	0,375	0,000	0,000	0,000
0,400	0,333	0,250	0,000	0,500	0,000	1,000	0,000
0,600	0,333	0,250	0,400	0,375	0,000	1,000	0,000
0,800	0,333	1,000	1,000	1,000	1,000	0,000	1,000
0,000	0,333	0,000	0,000	0,375	0,333	1,000	0,000
0,800	0,667	0,250	0,400	0,375	0,000	1,000	0,000
0,000	0,000	0,000	0,000	0,000	0,167	1,000	0,000
0,800	0,667	0,000	0,000	0,375	0,333	1,000	0,000
1,000	1,000	0,500	0,400	0,375	0,167	0,000	0,000
1,000	0,667	0,750	0,200	0,125	0,000	0,000	0,000

Randomly select the initial medoid as much as k from n data. The selected data are data no. 13, 14, and 15. It's shown in Tables 8 and 9.

TABLE VIII
INITIAL MEDOID DATA

Audiogram (db A) Left						Left Interpretation	
Degree Evaluation (Hz)			Preventive Evaluation (Hz)			Low Frequency	High Frequency
500	1000	2000	3000	4000	6000		
0,286	0,200	0,000	0,429	0,444	0,250	1,000	0,000
1,000	1,000	1,000	0,286	0,222	0,500	0,000	0,000
0,714	0,800	0,667	0,286	0,222	0,250	0,000	0,000

TABLE IX
INITIAL MEDOID ADVANCED DATA

Audiogram (db A) Right						Right Interpretation	
Degree Evaluation (Hz)			Preventive Evaluation (Hz)			Low Frequency	High Frequency
500	1000	2000	3000	4000	6000		
0,800	0,667	0,000	0,000	0,375	0,333	1,000	0,000
1,000	1,000	0,500	0,400	0,375	0,167	0,000	0,000
1,000	0,667	0,750	0,200	0,125	0,000	0,000	0,000

Here, the Euclidean distance formula is applied to calculate the length between each object to the transient medoid, signals the closest distance of the object to the medoid, and calculates the total by looking up its minimum and determining the cluster members against the transient medoid with the Normal Hearing Ability being cluster 1, the Mild Hearing Loss being cluster 2 and the Moderate Hearing Loss being cluster 3. Then, with the original medoid, the distance between one object and another should be calculated using the formula below. The result of the calculation is shown in Table 10.

$$d_{euc} = \sqrt{\begin{pmatrix} (0,286 - 0,429)^2 + (0,200 - 0,600)^2 + (0,000 - 0,500)^2 + \\ (0,429 - 0,286)^2 + (0,444 - 0,111)^2 + (0,250 - 0,000)^2 + \\ (1,000 - 1,000)^2 + (0,000 - 0,000)^2 + (0,800 - 0,800)^2 + \\ (0,667 - 0,667)^2 + (0,000 - 0,250)^2 + (0,000 - 0,000)^2 + \\ (0,375 - 0,125)^2 + (0,333 - 0,000)^2 + (1,000 - 1,000)^2 + \\ (0,000 - 0,000)^2 \end{pmatrix}} = 0,927652$$

TABLE X
CALCULATION OF DISTANCES AND INITIAL MEDOID CLUSTERS

No. 13	No. 14	No. 15	Proximity	Cluster
0,927652	1,860851	1,598204	0,927652	1
0,85352	2,146097	1,791425	0,85352	1
1,001657	2,181811	1,916989	1,001657	1
0,847417	2,027511	1,782174	0,847417	1
0,860072	1,904045	1,633982	0,860072	1
1,801552	1,905547	1,795711	1,795711	3
0,859364	2,247336	1,932721	0,859364	1
0,814742	2,045177	1,768269	0,814742	1
2,761937	2,74335	2,764155	2,74335	2
0,985719	2,271222	2,07162	0,985719	1
0,892489	1,884122	1,669076	0,892489	1
1,330905	2,774395	2,414878	1,330905	1
0	2,207121	1,968253	0	1
2,207121	0	0,774037	0	2
1,968253	0,774037	0	0	3
Total			13,9126	

Performs medoid iterations using data no. 1, 2, and 3 are shown in Tables 11 and 12.

TABLE XI
ITERATION DATA

Audiogram (db A) Left						Left Interpretation	
Degree Evaluation (Hz)			Preventive Evaluation (Hz)			Low Frequency	High Frequency
500	1000	2000	3000	4000	6000		
0,429	0,600	0,500	0,286	0,111	0,000	1,000	0,000
0,286	0,400	0,167	0,000	0,000	0,125	1,000	0,000
0,429	0,400	0,167	0,000	0,111	0,125	1,000	0,000

TABLE XII
ADVANCED DATA MEDOID ITERATION

Audiogram (db A) Right						Right Interpretation	
Degree Evaluation (Hz)			Preventive Evaluation (Hz)			Low Frequency	High Frequency
500	1000	2000	3000	4000	6000		
0,800	0,667	0,250	0,000	0,125	0,000	1,000	0,000
0,800	0,333	0,250	0,000	0,125	0,167	1,000	0,000
0,400	0,333	0,000	0,400	0,125	0,000	1,000	0,000

Recalculate the distance from each object at the 2nd iteration by using a new medoid, as shown in Table 13.

$$d_{euc} = \sqrt{\begin{pmatrix} (0,429 - 0,429)^2 + (0,600 - 0,600)^2 + (0,000 - 0,500)^2 + \\ (0,286 - 0,286)^2 + (0,111 - 0,111)^2 + (0,000 - 0,000)^2 + \\ (1,000 - 1,000)^2 + (0,000 - 0,000)^2 + (0,800 - 0,800)^2 + \\ (0,667 - 0,667)^2 + (0,250 - 0,250)^2 + (0,000 - 0,000)^2 + \\ (0,125 - 0,125)^2 + (0,000 - 0,000)^2 + (1,000 - 1,000)^2 + \\ (0,000 - 0,000)^2 \end{pmatrix}} = 0$$

TABLE XIII
CALCULATION OF DISTANCES AND INITIAL MEDOID CLUSTERS

No. 1	No. 2	No. 3	Proximity	Cluster
0	0,648083	0,861383	0	1
0,648083	0	0,665606	0	2
0,861383	0,665606	0	0	3
0,825932	0,600955	0,850027	0,600955	2
0,43211	0,558841	0,77036	0,43211	1
1,846539	1,995679	2,019021	1,846539	1
0,854123	0,697409	0,738437	0,697409	2
0,739738	0,782841	0,668014	0,668014	3
2,971153	3,006359	3,018573	2,971153	1
1,109335	1,078701	0,884473	0,884473	3
0,621911	0,753948	0,710562	0,621911	1
1,489505	1,077565	0,959589	0,959589	3
0,927652	0,85352	1,001657	0,85352	2
1,860851	2,146097	2,181811	1,860851	1
1,598204	1,791425	1,916989	1,598204	1
Total			13,99473	

Calculating the total deviation is by looking for the difference between the total or amount of new and old proximity. The difference obtained between the total proximity of the new and old medoids is 0.082127799. Since the proximity cell ion is larger than 0, the iteration stops, and the cluster exists in the previous iteration. In the first iteration, the cluster obtained was cluster 1 with a total of 11 people, cluster 2 with two people, and cluster 3 with two people.

TABLE XIV
CLUSTER RESULTS ON THE INITIAL MEDOID

No. 13	No. 14	No. 15	Proximity	Cluster
0,927652	1,860851	1,598204	0,927652	1
0,85352	2,146097	1,791425	0,85352	1
1,001657	2,181811	1,916989	1,001657	1
0,847417	2,027511	1,782174	0,847417	1
0,860072	1,904045	1,633982	0,860072	1
1,801552	1,905547	1,795711	1,795711	3
0,859364	2,247336	1,932721	0,859364	1
0,814742	2,045177	1,768269	0,814742	1
2,761937	2,74335	2,764155	2,74335	2
0,985719	2,271222	2,07162	0,985719	1
0,892489	1,884122	1,669076	0,892489	1
1,330905	2,774395	2,414878	1,330905	1
0	2,207121	1,968253	0	1
2,207121	0	0,774037	0	2
1,968253	0,774037	0	0	3
Total			13,9126	

D. *Testing K-Medoids Method with RapidMiner*

K-Medoids algorithm testing was carried out by comparing the results of K-Medoids processing with systems built with K-Medoids processing through the RapidMiner application. Below is the comparison of the data.

Figures 5 and 6 show the difference between the results obtained using the RapidMiner application and the system

built with cluster 0 (normal hearing ability category), cluster 1 (mild hearing loss category), and cluster 2 (moderate hearing loss category).

1) SMA 1 Padang

Cluster Model

```
Cluster 0: 11 items
Cluster 1: 1 items
Cluster 2: 3 items
Total number of items: 15
```

Fig. 5 Rapid Miner clustering results of SMAN 1 Padang

Student Data categorized as Normal Hearing Ability

Search ..   Search

Total students   : 9 persons
Female students   : 8 persons
Male students   : 1 persons

Student Data categorized as Mild Hearing Loss

Search ..   Search

Total students   : 2 persons
Female students   : 1 persons
Male students   : 1 persons

Student Data categorized as Moderate Hearing Loss

Search ..   Search

Total students   : 4 persons
Female students   : 4 persons
Male students   : 0 persons

Fig. 6 System clustering results of SMAN 1 Padang

2) SMA 5 Padang

Cluster Model

```
Cluster 0: 11 items
Cluster 1: 1 items
Cluster 2: 3 items
Total number of items: 15
```

Fig. 7 Rapid Miner clustering results of SMAN 5 Padang

Student Data categorized as Normal Hearing Ability

Search ..   Search

Total students   : 2 persons
Female students   : 1 persons
Male students   : 1 persons

Student Data categorized as Mild Hearing Loss

Search ..   Search

Total students   : 8 persons
Female students   : 5 persons
Male students   : 3 persons

Student Data categorized as Moderate Hearing Loss

Search ..   Search

Total students   : 6 persons
Female students   : 4 persons
Male students   : 2 persons

Fig. 8 System clustering results of SMAN 5 Padang

Figs. 7 and 8 show the difference between the results obtained using the RapidMiner application and the system built with cluster 0 (normal hearing ability category), cluster 1 (mild hearing loss category), and cluster 2 (moderate hearing loss category).

3) SMA 6 Padang

Cluster Model

```
Cluster 0: 11 items
Cluster 1: 1 items
Cluster 2: 3 items
Total number of items: 15
```

Fig. 9 Rapid Miner clustering results of SMAN 6 Padang

Student Data categorized as Normal Hearing Ability

Search ..   Search

Total students   : 15 persons
Female students   : 11 persons
Male students   : 4 persons

Student Data categorized as Mild Hearing Loss

Search ..   Search

Total students   : 4 persons
Female students   : 4 persons
Male students   : 0 persons

Student Data categorized as Moderate Hearing Loss

Search ..   Search

Total students   : 3 persons
Female students   : 2 persons
Male students   : 1 persons

Fig. 10 System clustering results of SMAN 6 Padang

The figures above show a difference between the results obtained using the RapidMiner application and the system built with cluster 0 (normal hearing ability category), cluster 1 (mild hearing loss category), and cluster 2 (moderate hearing loss category).

E. Confusion Matrix Testing

Based on tests conducted using the K-Medoids method, the confusion matrix is employed to compute the method's performance. In the data source, there are left and right interpretations. The interpretation has values of R (Normal Hearing Ability), N (Mild Hearing Loss), and T (Moderate Hearing Loss), where the values of R and T are initialized to be negative and the value of N is initialized to be positive in the Confusion Matrix. Likewise, in the clustering data, categories R (Normal), N (Mild), and T (Moderate), where categories R and T are initialized to be negative and category N is initialized to be positive in the Confusion Matrix, as seen in Table 15.

TABLE XV
CONFUSION MATRIX MEASUREMENTS

Total Students: 53		K-Medoids algorithm	
		True	False
Original Data	True	9	33
	False	5	6

Below is the result of calculating the accuracy value.

$$accuracy = \left(\frac{9+6}{9+33+5+6}\right) = 0.283$$
$$accuracy = 0.283 \times 100\%$$
$$accuracy = 28.3\%$$

Below is the result of calculating the precision value.

$$precision = \left(\frac{9}{9+5}\right) = 0.643$$
$$precision = 0.643 \times 100\%$$
$$precision = 64.3\%$$

Below is the result of calculating the recall value.

$$recall = \left(\frac{9}{9+33}\right) = 0.214$$
$$recall = 0.214 \times 100\%$$
$$recall = 21.4\%$$

Using the confusion matrix method, an accuracy value of 28.3%, a precision value of 64.3%, and a recall value of 21.4% were obtained.

IV. Conclusion

This research found that with the K-Medoids system, schools can see student data that has an impact on the ears. The K-Medoids method can be applied in this system to determine the type of students' ears. The results of clustering depend on the comparison of the difference between the data that is iterated and the medoid data that has been selected, provided that the data to be iterated is no longer derived from the medoid data that has been selected for the first time. From the data of several high school students in Padang, none of them are included in clusters 4 (moderately severe hearing loss), 5 (severe hearing loss), and 6 (profound hearing loss), meaning that none of the students have very serious ear disorders. In the data, students from SMA 3, SMA 8, and SMA 10 cannot cluster because the minimum and maximum values were the same at the time the data was transformed. So, it must be normalized and continued.

References

[1] J. H. V. Caracol, B. Alturas, and A. Martins, "A society ruled by the impact of the smartphone: Influence that the use of the smartphone has in people's daily lives," 2019 14th Iberian Conference on Information Systems and Technologies (CISTI), Jun. 2019, doi:10.23919/cisti.2019.8760845.

[2] G.-J. Wang et al., "Technical Report: A New Device Attached to a Smartphone for Objective Vision Screening," Optometry and Vision Science, vol. 98, no. 1, pp. 18–23, Jan. 2021, doi: 10.1097/opx.0000000000001621.

[3] M. Gertz, S. Schütz-Bosbach, and S. Diefenbach, "Smartphone and the Self: Experimental Investigation of Self-Incorporation of and Attachment to Smartphones," Multimodal Technologies and Interaction, vol. 5, no. 11, p. 67, Oct. 2021, doi: 10.3390/mti5110067.

[4] В. П. Шейнов and Н. В. Дятчик, "Psychological States And Properties Of The Personality Having Dependence On The Smartphone," Pedagogical Review, no. 6(40), pp. 225–235, Dec. 2021, doi: 10.23951/2307-6127-2021-6-225-235.

[5] S. Kumari, "A study of the impact of electronic devices on the lifestyle of secondary school students.," International Journal Of Humanities, Engineering, Science And Management, vol. 4, no. 1, pp. 01–07, Jun. 2023, doi: 10.59364/ijhesm.v4i1.228.

[6] A. S. Mathias Duarte et al., "Audiogram Comparison of Workers from Five Professional Categories," BioMed Research International, vol. 2015, pp. 1–5, 2015, doi: 10.1155/2015/201494.

[7] Z. Musiba, "Classification of audiograms in the prevention of noise-induced hearing loss: A clinical perspective," South African Journal of Communication Disorders, vol. 67, no. 2, Mar. 2020, doi:10.4102/sajcd.v67i2.691.

[8] R. Guo, R. Liang, Q. Wang, and C. Zou, "Hearing loss classification algorithm based on the insertion gain of hearing aid," Multimedia Tools and Applications, vol. 82, no. 26, pp. 41225–41239, Apr. 2023, doi: 10.1007/s11042-023-14886-0.

[9] T. Li, M. Zhang, Y. Li, E. Lagerspetz, S. Tarkoma, and P. Hui, "The Impact of Covid-19 on Smartphone Usage," IEEE Internet of Things Journal, vol. 8, no. 23, pp. 16723–16733, Dec. 2021, doi:10.1109/jiot.2021.3073864.

[10] E. Łuszczki et al., "Children's Eating Habits, Physical Activity, Sleep, and Media Usage before and during COVID-19 Pandemic in Poland," Nutrients, vol. 13, no. 7, p. 2447, Jul. 2021, doi: 10.3390/nu13072447.

[11] S. E. Narayanan and P. Manjula, "Effect of Compression Type and Number of Channels on Speech Perception in Quiet in Listeners with Different Audiogram Configurations," Auditory and Vestibular Research, Jun. 2022, doi: 10.18502/avr.v31i3.9872.

[12] M. Ohsaki, P. Wang, K. Matsuda, S. Katagiri, H. Watanabe, and A. Ralescu, "Confusion-Matrix-Based Kernel Logistic Regression for Imbalanced Data Classification," IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 9, pp. 1806–1819, Sep. 2017, doi:10.1109/tkde.2017.2682249.

[13] E. Herman, K.-E. Zsido, and V. Fenyves, "Cluster Analysis with K-Mean versus K-Medoid in Financial Performance Evaluation," Applied Sciences, vol. 12, no. 16, p. 7985, Aug. 2022, doi: 10.3390/app12167985.

[14] M. Nigar Shumaila, "A Comparison of K-Means and Mean Shift Algorithms," International Journal of Theoretical and Applied Mathematics, vol. 7, no. 5, p. 76, 2021, doi:10.11648/j.ijtam.20210705.12.

[15] N. K. Kaur, U. Kaur, and D. Singh, "K-Medoid Clustering Algorithm-A Review K-Medoid Clustering Algorithm- A Review."

[16] S. Mousavi, F. Z. Boroujeni, and S. Aryanmehr, "Improving customer clustering by optimal selection of cluster centroids in K-means and K-medoids algorithms," J. Theor. Appl. Inf. Technol., vol. 8, no. 10, 2020.

[17] P. K. Bhuyan and K. V. K. Rao, "Defining Los Criteria of Urban Streets Using GPS Data: K-Means and K-Medoid Clustering in Indian Context," TRANSPORT, vol. 27, no. 2, pp. 149–157, Jun. 2012, doi:10.3846/16484142.2012.692354.

[18] A. Bhat, "K-Medoids Clustering Using Partitioning Around Medoids for Performing Face Recognition," International Journal of Soft Computing, Mathematics and Control, vol. 3, no. 3, pp. 1–12, Aug. 2014, doi: 10.14810/ijscmc.2014.3301.

[19] Velmurugan, "Computational Complexity between K-Means and K-Medoids Clustering Algorithms for Normal and Uniform Distributions of Data Points," Journal of Computer Science, vol. 6, no. 3, pp. 363–368, Mar. 2010, doi: 10.3844/jcssp.2010.363.368.

[20] R. Alqaisi, W. Ghanem, and A. Qaroush, "Extractive Multi-Document Arabic Text Summarization Using Evolutionary Multi-Objective Optimization With K-Medoid Clustering," IEEE Access, vol. 8, pp. 228206–228224, 2020, doi: 10.1109/access.2020.3046494.

[21] I. Markoulidakis, G. Kopsiaftis, I. Rallis, and I. Georgoulas, "Multi-Class Confusion Matrix Reduction method and its application on Net Promoter Score classification problem," Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference, Jun. 2021, doi: 10.1145/3453892.3461323.

[22] C. S.Dangare and S. S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques," International Journal of Computer Applications, vol. 47, no. 10, pp. 44–48, Jun. 2012, doi: 10.5120/7228-0076.

[23] S. Daneshtalab and H. Rastiveis, "Decision Level Fusion of Orthophoto And Lidar Data Using Confusion Matrix Information for LNAD Cover Classification," The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLII-4/W4, pp. 59–64, Sep. 2017, doi: 10.5194/isprs-archives-xlii-4-w4-59-2017.

[24] V. M. Patro and M. Ranjan Patra, "Augmenting Weighted Average with Confusion Matrix to Enhance Classification Accuracy," Transactions on Machine Learning and Artificial Intelligence, vol. 2, no. 4, Aug. 2014, doi: 10.14738/tmlai.24.328.

[25] S. Defiyanti, M. Jajuli, and N. Rohmawati, "K-Medoid Algorithm in Clustering Student Scholarship Applicants," Scientific Journal of Informatics, vol. 4, no. 1, pp. 27–33, May 2017, doi:10.15294/sji.v4i1.8212.

[26] P. Arora, Deepali, and S. Varshney, "Analysis of K-Means and K-Medoids Algorithm For Big Data," Procedia Computer Science, vol. 78, pp. 507–512, 2016, doi: 10.1016/j.procs.2016.02.095.

[27] P. Breiding, F. Sottile, and J. Woodcock, "Euclidean Distance Degree and Mixed Volume," Foundations of Computational Mathematics, vol. 22, no. 6, pp. 1743–1765, Sep. 2021, doi: 10.1007/s10208-021-09534-8.

[28] F. Rahmad, Y. Suryanto, and K. Ramli, "Performance Comparison of Anti-Spam Technology Using Confusion Matrix Classification," IOP Conference Series: Materials Science and Engineering, vol. 879, no. 1, p. 012076, Jul. 2020, doi: 10.1088/1757-899x/879/1/012076.

[29] M. Heydarian, T. E. Doyle, and R. Samavi, "MLCM: Multi-Label Confusion Matrix," IEEE Access, vol. 10, pp. 19083–19095, 2022, doi:10.1109/access.2022.3151048.

[30] S.-C. Kim and Y.-S. Cho, "Predictive System Implementation to Improve the Accuracy of Urine Self-Diagnosis with Smartphones: Application of a Confusion Matrix-Based Learning Model through RGB Semiquantitative Analysis," Sensors, vol. 22, no. 14, p. 5445, Jul. 2022, doi: 10.3390/s22145445.