Building Dental Identification System Using the Apriori Algorithm via Haar Filter.

Ass.lecturer Ziadoon W. Salman
Ministry of Education Directorate of Education Rusafa – 2

بناء نظام التعرف على الأسنان باستخدام خوارزمية أبريوري عبر فلتر هار

م.م زيدون وليد سلمان
مديرية تربية بغداد الرصافة الثانية

Abstract

A dental identification system uses a digitized dental image as the foundation for identifying people. The dental identification systems offer an automatic matching by looking for similarities in photographic dental images. The proposed DIS approach, which is discussed in this research, largely improves the dental identification technique using the Apriori algorithm, with the goal of resolving the drawbacks of dental rule mining. In this instance, the association rules technique was employed to extract the relationships between the wavelet-based Haar filter-extracted dental image characteristics. Using a collected database of 63 dental images, the suggested DIS was put to the test. The suggested approach uses one, two, and then three samples for training and testing, although we have found that training three samples yields the best results. where three samples yielded values that were 90%, 70% from two samples, and 60% from one sample.

المستخلص

يستخدم نظام التعرف على الأسنان صورة الأسنان الرقمية كأساس لتحديد هوية الأشخاص. توفر أنظمة تحديد الأسنان مطابقة تلقائية من خلال البحث عن أوجه التشابه في صور الأسنان الفوتوغرافية. إن نهج DIS المقترح ، الذي تمت مناقشته في هذا البحث ، يحسن إلى حد كبير تقنية تحديد الأسنان باستخدام خوارزمية Apriori ، بهدف حل عيوب التنقيب عن قواعد الأسنان. في هذه الحالة ، تم استخدام تقنية قواعد

الارتباط لاستخراج العلاقات في الصور بواسطة الموجة ثنائية الأبعاد عبر خصائص صورة الأسنان المستخرجة بواسطة مرشح الـ Haar. باستخدام قاعدة بيانات مجمعة من ٦٣ صورة أسنان ، تم اختبار DIS المقترحة. الأسلوب المقترح يستخدم عينة واحدة ، اثنتين ، ثم ثلاث عينات للتدريب والاختبار ، وجدنا أن تدريب ثلاث عينات يحقق أفضل النتائج. حيث أعطت ثلاث عينات قيم كانت ٩٠٪ ، ٧٠٪ من عينتين ، و ٦٠٪ من عينة واحدة.

KEYWORDS: Dental Identification System, Association Rule, Apriori Algorithm, Haar Filter, Discrete Wavelet Transform.

I.INTRODUCTION

Data mining (DM) has gained popularity across all academic disciplines (from business to science) and the emergence of systems utilizing biometrics is no exception. The use of knowledge discovery techniques to biometric information in order to identify implicit patterns is known as "biometric data mining" (BDM). Finding characteristics of case subsets that differ from the other instances is the primary objective of the majority of data mining problems in the biometrics research area [1]. Through the recent evolution of information technology, biometric technologies have been quickly expanding in the field of recognizing and analyzing identification. The face, voice, fingerprint, signature, iris, and tooth are only a few examples of the physiological or behavioral characteristics used in these systems to identify people [2]. One of the best biometrics for monitoring purposes is dental. Because dental recognition technology may be used in a variety of environments, including access control, identity authentication, and other uses, dentistry has seen an increase in interest in computer vision and machine learning over the past several years. The process of matching a dental image to any of them is known as dental recognition. Although there has been great progress made in the study of how to recognize teeth despite little variations in shape and illumination, the present systems are still a long way from being able to match the individual perception system [3].

II. RELATED WORK

- In **[Gowri Vijay Reesu, 2020],** Employing a dental identification scenario, a unique automated human dental identification technique is presented. 120 intra-oral scans-IOS and 120 dental models were divided into two groups, were created from the entire research sample. 30 maxillaries and 30 mandibular 3D-scanned dental models of individuals who had received orthodontic treatment made up Group A data. This information was regarded as digital AM data. For the purpose of producing dental casts (60) of the same people who made the same sample were obtained. The success percentages for the maxillary and mandibular IOS ranged from (64) to (100% and 81 to 100%, respectively, with an overall average of 96.7 and 96.4) [4].
- In **[Gowri Vijay Reesu, 2020],** This study seeks to examine cutting-edge odonatological approaches by fusing 2D images with 3D dental modeling and dental identity scenarios. By integrating an AM image with imaging technologies as an alternative to postpartum (PM) photographs, The objective of this study was to improve dental identification accuracy. The study employed 35 digital images to mimic AM information and 31 3D dental models to simulate PM data. The obtained data underwent two phases of analysis: Phase 2: 2D-3D super imposition after a wash delay, after Stage 1: ocular examination of 2D-3D images. [5].

- In **[Qingsong,2021],** This paper introduces a novel concept called A human identification network called the attention-based multisupervision network (AMNet). In order to combine multilevel global features, AMNet contains feature fusion branches, part feature branches to produce discriminative local features, and a strategy to monitoring focused on attention to obtain an appropriate attention mask. Following the extraction of features from dental panoramic X-ray pictures, cosine similarity is used to generate matching scores between query and gallery features to ascertain if the identity shared by the query image and gallery image. The query set, which included 665 dental panoramic X-ray images from 503 distinct participants, was accurately answered by the proposed approach with rank-one accuracy of 88.72% and rank-five accuracy of 95.79% [6].
- In **[Hideko Fujimoto,2021],** Using head and neck CT and MR imaging to acquire identifying information, including dental results, this study sought to identify people. Resources and Techniques We examined dental outcomes and dental arrangement using head and neck images, and we searched for morphological similarities by superimposing. Additionally, a 3D superimposition using cone-beam CT imaging from both the antemortem (AM) and post-mortem (PM) was carried out. Results: All instances matched in 3D when the superimposition of PM-CT, AM-CBCT, and AM-MR pictures was evaluated. Dental results were comparable when AM and PM pictures were compared and combined. Conclusion: The results indicate that, even in the absence of AM dental pictures, dental findings may be evaluated on head and neck images [7].

III. Haar Basis Filter

The following equations define the ((Haar basis filter) high pass filter (HPF)) and (low pass filter (LPF)), both of which [8]:

$$\text{HPF: } \frac{1}{\sqrt{2}}[1 \quad -1], \ldots\ldots (1)$$

$$\text{LPF: } \frac{1}{\sqrt{2}}[1 \quad 1], \ldots\ldots (2)$$

The filters (HP, LP) are referred to as "decomposition filters" since they divide up or break down the image into approximate and detailed coefficients, respectively. By convolution with the band-pass filter in a given direction, a referred to as details image is created, and by convolution with the LPF, a referred to as approximation image is created. Applying LPF in both horizontal and vertical directions results in the LL band (approximation band), and its filter is constructed as follows [8]:

$$LL = \frac{1}{\sqrt{2}}(1 \quad 1)^t . \frac{1}{\sqrt{2}}(1 \quad 1) = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \ldots\ldots (3)$$

Using vertical HPF and horizontal LPF results in the LH band (detail band), and its filter is produced as follows:

$$LH = \frac{1}{\sqrt{2}}(1 \quad 1)^t . \frac{1}{\sqrt{2}} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}, \ldots\ldots (4)$$

Applying horizontal HPF and vertical LPF results in the HL band (detail band), and its filter is produced as follows:

$$HL = \frac{1}{\sqrt{2}}(1 \quad -1)^t . \frac{1}{\sqrt{2}}(1 \quad 1) = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}, \ldots\ldots (5)$$

The utilization of horizontal and vertical HPF results in the HH band (detail band) and its filter is produced as follows:

$$HH = \frac{1}{\sqrt{2}}(1 \quad -1)^t . \frac{1}{\sqrt{2}}(1 \quad -1) = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \dots\dots\dots\dots\dots \quad \dots. \quad (6)$$

By applying the previously discussed filters to each (2*2) neighboring pixel of the whole image, modified bands (LL scaling bands and three wavelet bands HL, LH, and HH) may be made [9].Below is an illustration of how to compute the Haar wavelet transform for N items such as: Calculate the N/2 average for each pair of samples [10].

- Determine the variance in each average and sample. By utilizing (N/2 differences), it is calculated.
- Place the averages in the array's first half.
- Insert the differences into the array's second half.
- Repeat step 5 on the array's first half, (The array's length must be a power of 2).

IV. Decomposition of Discrete Wavelet Transform (DWT)

The DWT core technique is multi-resolution analysis (MRA), is an idea. Comparing the DWT of image signals to other multiscale representations like the Laplacian and Gaussian pyramid, with more spectral and geographical localisation of image production, the DWT generates a non-redundant image representation.. More and more people are becoming interested in image denoising lately thanks to the discrete wavelet transform. When two complementary filters are applied to the signal S, two signals—an approximation and a detail are created. This is a decomposition or analysis. When the pieces are placed back together to reconstruct the original signal, there won't be any information loss. Reconstruction or synthesis is the term used to describe this process. Analysis and synthesis are involved in the mathematical processes of a DWT and an inverse DWT [11].

The DWT for two dimensional images x[m,n] may be defined identically as follows by applying one dimensional DWT to each dimension m and n separately:

N= [DWT_m (x[m,n])]. An imagine is divided into "sub-bands" by 2-dimensional WT that are uniform in direction and frequency. WT is produced by running the image through a number of filter bank stages. when an image initially has a horizontal filter applied to it [9]. Examples of limited pulsation reply filters are scaling function LPF and wavelet function HPF. In other words, just a little portion of the input is used to determine the output at each stage. The outputs of the filters are then horizontally downsampled by a ratio of 2. Then, using two similar filters, each of those signals is vertically filtered. Decompose the picture eventually into the 4 subbands denoted by LL, HL, LH, and HH. You may think of each of the sub-bands as a scaled-down representation of the image that depicts a different aspect of the image. A rougher approximation of the original image is the Low-Low. Low High & High Low are employed successively to record the variations in the image's vertical and horizontal axes. High provides the image's high-frequency element. The Low Low sub-band may then be decomposed into two levels [12].

V. The Apriori Algorithm

A important data mining technique is the Apriori algorithm that is used to produce association rules. One of the approaches used in association rules (AR) has an impact on the frequency item sets. The apriori algorithm is a set of procedures that must be followed in order to identify the most frequent itemset in a certain database. The join and trim processes are repeated in this data mining method until the most frequent itemset is created [13]. A minimal support threshold must be specified by the problem for the user to avoid assuming it. The idea was put out by Srikant and Agrawal. The method is used to build the association rules and to mine frequently occurring itemsets. The basic principle of the method is that every subset of a frequent itemset is also a frequent itemset. By only examining the item sets with a support count higher than the minimal support count, it limits the pool of candidates who are taken into consideration. If a collection of objects has an uncommon subset, all infrequent itemsets may be trimmed. The approach performs a level-wise search to investigate "(k+1)-item sets" using "k-item sets". The basic apriori algorithm's stages are illustrated in algorithm (1). This method uses the "join" and "prune" stages to reduce the search space. Iterative search is used to determine the most frequent item sets. If there is a chance that item I is unlikely to be typical [14].

Algorithm (1) Basic Apriori algorithm
Input: Database D, of the transaction; minimum support; Threshold; min-sup in D.
Output: frequent item sets in D.

Step 1: In the first iteration of the algorithm, each item is taken as a 1-itemsets candidate. The algorithm will count the occurrences of each item.

Step 2: Let there be some minimum support, min_sup. The set of 1-item sets whose occurrence is satisfying the min sup is determined. Only those candidates which count more than or equal to min_sup, are taken ahead for the next iteration and the others are pruned.

Step 3: Next, 2-itemset frequent items with min_sup are discovered. For this in the join step, the 2-itemset is generated by forming a group of 2 by combining items with itself.

Step 4: The 2-itemset candidates are pruned using a min-sup threshold value. Now the table will have 2 –item sets with min-sup only.

Step 5: The next iteration will form 3 –item sets using the join and prune step. This iteration will follow the ant monotone property where the subsets of 3-itemsets that is the 2 –item set subsets of each group fall in min_sup. If all 2-itemset subsets are frequent then the superset will be frequent otherwise, it is pruned.

VI. The Proposed System

The core idea behind the proposed DIS approach is based on the fact that every dental in every human dental image has a different set of characteristics. These qualities are different for each human dental imaging. In this research, methods for the DIS system are suggested to extract distinctive dental characteristics from different dental images, and to recognize the input image of a person's mouth by learning association rules between these elements. Discrete wavelet transform (DWT)-based Haar is used to extract image texture data during the training phase. Each human dental image is analyzed using a (sub-band) of filtered wavelet coefficient images. The wavelet coefficients are used to depict the dental texture. In order to represent every image of individuals with dentition in data mining, association rules are then created using the recovered textural elements. In the testing stage of the proposed IDS, all the features amassed during the training phase are kept in the features database and used to locate the nearest human dental picture using the association rules algorithm's minimum distance matching. As illustrated in Figure (2), the proposed dental identification system's structure, The preprocessing and feature extraction methods used in both phases are the same. The following are the stages for the suggested DIS:

• As component of the pre-processing on the input images, execute image cropping and contrast enhancing to define the area of interest (ROI) for the feature extraction step.

• Extracting features from dental images.

• Produce a feature vector.

• Use the Apriori algorithm for association rule mining.

• Removing the association's regulations.

• Decisions result.

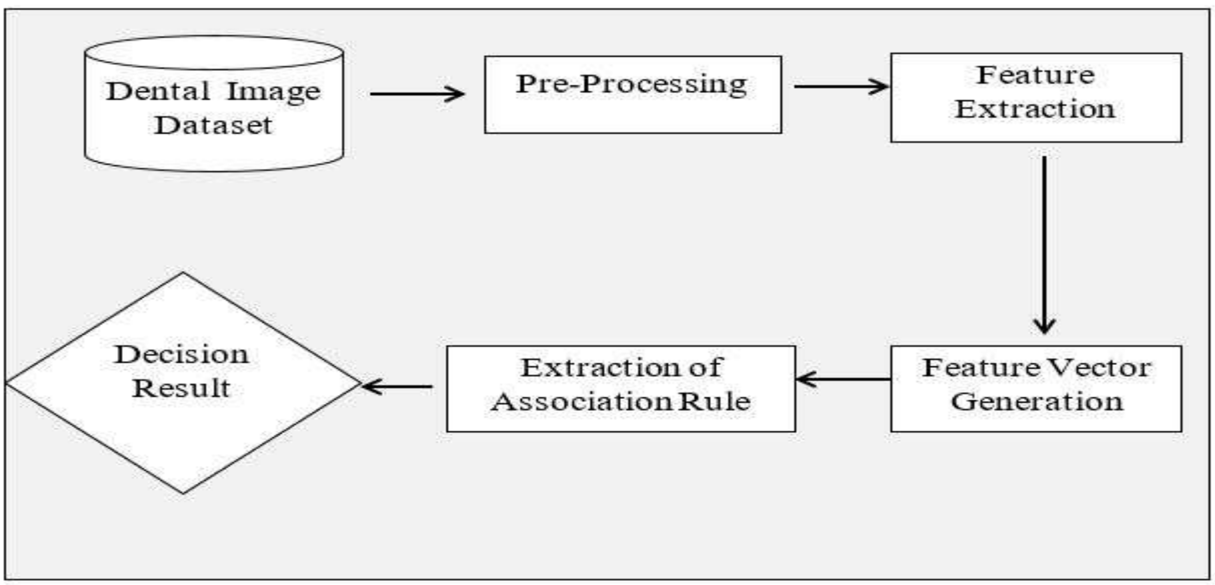

Figure (2): The proposed dental identification technique's framework.

A. pre-processing Image

Each dental image used in the present research has been enhanced and cropped to lessen data volume, which quickens processing while improving feature extraction precision. In this step of dental imaging, contrast image improvement is achieved using histogram equalization. A nonlinear process called histogram equalization

(HE) aims to highlight image brightness in a way that is compatible with the human visual system. The HE algorithm aims to improve the quality of an image to produce on … er of … e

$$I(X, Y) = 0.99R(X, Y) + 0.587G(X, Y) + 0.114B(X, Y) \quad \ldots\ldots\ldots\ldots \quad (7)$$

histogram is just a list of each pixel value in a picture. The developed procedures for dental detection pre-processing are shown in Algorithm (2). To achieve the best results, a color dental picture is converted into a grayscale image during the pre-processing step using the weighted technique as in equation (7).

Where I(X, Y) is the grayscale image, R, G, and B are the three color space images.

Algorithm (2) The Histogram equalization (HE) algorithm
Input: Original image
Output: image cropping (dental cropping)

Step1: Convert the input dental image into a greyscale image by using the equation (7).

Step2: Calculate the frequency of each pixel value occurrence for the input dental image.

Step3: Find the cumulative frequency of pixel value occurrence.

Step4: Divide the cumulative frequencies from step 3 by the total number of dental image pixels and then multiply by the maximum pixel value of the input dental image.

Step5: image cropping // preserving the important regions of the dental input image.
Step6: End.

B. Feature Extraction

A principal (feature extraction) of the proposed system used the Haar wavelet transform. The algorithm (3) steps are used to implement the feature extraction technique. The process of "feature extraction" entails breaking down the representation of a picture into a manageable amount of components that include sufficient discriminating information. Computing a collection of moments is an effective technique to characterize textural information and reduce dimensionality. They will calculate statistical metrics for each band LL2, HL2, HH2, LH2, and HH2 frequencies, are usually effective for highlighting the traits or aspects of data. In section III, adopted statistical measures are presented.

Algorithm (3) Haar 2D wavelet transform.
Input: image cropping
Output: Extract texture information (contrast, correlation, energy, and homogeneity)

Step1: Read dental image data
Step2: Detection process for the dental area by using the algorithm (2)
Step3: Apply the 2D wavelet decomposition using a Haar filter wavelet. Then calculate the "approximation coefficients" matrix (AC) with the "detail coefficient" matrixes of horizontal, vertical, and diagonal. // The Haar Basis filter is performed by using the equations that are mentioned in section III.
Step4: Calculation of contrast, correlation, energy, and homogeneity features, for each input image, which are, represented the approximation coefficients matrix (AC) at level 2 of Haar wavelet.
Step5: Store the extracted features in the training database.
Step6: Repeating the steps from (1-6) for all input dental images.
Step7: End.

C. Association Rules Mining (AR)

From the dental images' AR that were extracted. AR mining looks through an enormous amount of images to identify the most notable relationships by emphasizing the characteristics that are commonly present in each dataset. "AR mining" is the process of locating often recurring item sets that provide robust association rules that satisfy the minimum confidence requirement. The association rule extraction process could be finished with the support of the Apriori algorithm. Apriori is an effective technique for often occurring itemets. While Apriori is based on minimal quantities support (MQS) and quantity confidence, this depends on numerous frequencies and confidence. The Association Rules (AR) algorithm's steps are shown in Algorithm (4). The section of transactions known as "Confidence" is made up of things from B that may be found in transactions that also contain a: Confidence (A, B) equals P (B|A). Equations (8) and (9) that assess support and

cc

Support (A→ B) = (Number of tuples (transactions) containing both X and Y)/ (Total number of tuples (transactions)) ……..… (8)

pr va

Confidence (A→ B) = (Number of tuples (transactions) both X and Y)/ (Total number of tuples (transactions) containing X) ……… (9)

The objective of mining AR is to find all rules with (Support and Confidence) values bigger than the minimal Support and minimum Confidence.

Algorithm(4) Implementation Association Rules Mining (Apriori algorithm) **Input:** Texture information features **Output:** Extract Association Rules
Step 1: Find out the support of the dental image database itemsets by using equation (8) **Step 2:** Choose the minimum value of both the support

The user's selected threshold levels. The transaction attributes of each input dental image are entered into the Apriori algorithm, which then generates the required association rules.

VII. EXPERIMENTAL RESULTS

A resulting identification rate for the training and testing samples is shown in Table (1) and Chart (1). When using (one, two, and three samples) training, the training with three samples yields the highest recognition rate. The proposed system's performance is measured using two metrics: False Alarm Rate (FAR), Recognition Rate (RR).

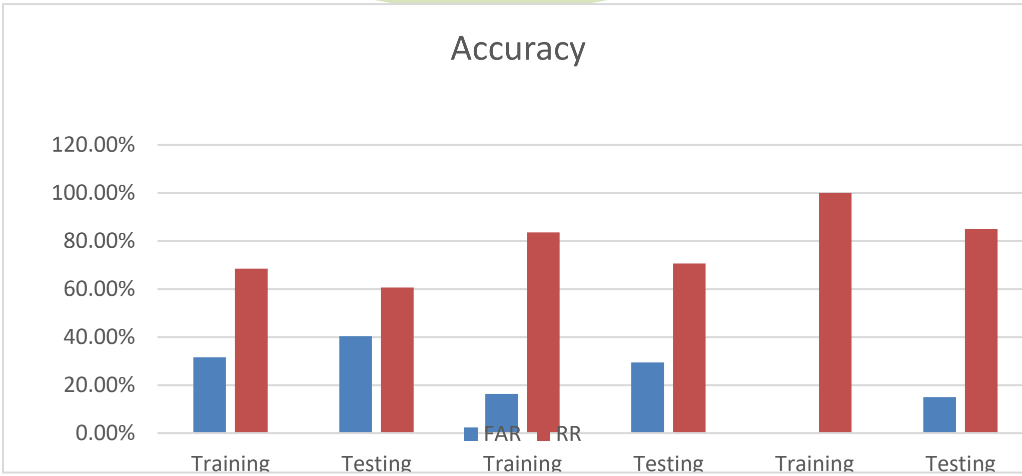

Table (1) the attained identification rate for the training and testing for all classes.

Identification Rate				
System Phase	Evaluation Criteria			
	Total Attempts	False Attempts	FAR	RR
Training	40	20	30%	70%
Testing	310	160	40%	60%
Training	95	15	20%	80%
Testing	300	90	30%	70%
Training	150	0	0.0	100%

Figure 1: Chart of Accuracy

The training of three samples produced the greatest results, with an identification accuracy for the test phase of 90%. Compared to training three samples, the outcomes from employing one or two training samples were unsatisfactory.

VIII. CONCLUSIONS

The proposed system's results have produced a number of conclusions, which may be summed up as follows:

- Using the Haar two-level discrete wavelet transform on the wavelet coefficients in combination with reliable statistical criteria like contrast, correlation, energy, and homogeneity yields the best results.
- Higher identification and retrieval performance outcomes are obtained when Association Rules are used to identify and recover dental images.
- The suggested approach employs one sample, two samples, and eventually three samples, but the greatest outcomes come from using just one sample.
- By applying the Haar two-level discrete wavelet transform to the wavelet coefficients and using reliable statistical criteria like contrast, correlation, energy, and homogeneity, the optimal outcome is attained.
- Using the Apriori algorithm, the method described in this thesis achieves a (90%) recognition rate for the Association Rules.

REFERENCES

[1] Albert, Midori, ed. *Biometrics: Unique and Diverse Applications in Nature, Science, and Technology*. BoD–Books on Demand, 2011.

[2] Guennouni, Souhail, Anass Mansouri, and Ali Ahaitouf. "Biometric systems and their applications." In *Visual impairment and blindness-what we know and what we have to know*. IntechOpen, 2019.

[3] Belhadj, Foudil. "Biometric system for identification and authentication." PhD diss., Ecole nationale Supérieure en Informatique Alger, 2017.

[4] Reesu, Gowri Vijay, Brenainn Woodsend, Scheila Mânica, Gavin F. Revie, Nathan L. Brown, and Peter A. Mossey. "Automated Identification from Dental Data (AutoIDD): A new development in digital forensics." *Forensic science international* 309 (2020): 110218.

[5] Jayakumar, Jayapriya, and Scheila Mânica. "The role of charting dental anomalies in human identification." *Forensic Science International: Reports* 2 (2020): 100086.

[6] Wu, Qingsong, Fei Fan, Peixi Liao, Yancun Lai, Wenchi Ke, Wenchao Du, Hu Chen, Zhenhua Deng, and Yi Zhang. "Human identification with dental panoramic images based on deep learning." *Sensing and Imaging* 22 (2021): 1-19.

[7] Fujimoto, Hideko. "Dental radiographic identification using ante-mortem CT, cone-beam CT, and MRI head and neck assessments." *Forensic Imaging* 26 (2021): 200465.

[8] Schimmack, Manuel, Susan Nguyen, and Paolo Mercorelli. "Anatomy of Haar wavelet filter and its implementation for signal processing." *IFAC-PapersOnLine* 49, no. 6 (2016): 99-104.

[9] Bnou, Khawla, Said Raghay, and Abdelilah Hakim. "A wavelet denoising approach based on unsupervised learning model." *EURASIP Journal on Advances in Signal Processing* 2020, no. 1 (2020): 1-26.

[10] Pasnur, M. A., and P. S. Malge. "Image retrieval using modified haar wavelet transform and K means clustering." *International Journal of Emerging Technology and Advanced Engineering* 3, no. 3 (2013): 89-93.

[11] Zhang, Bai-Ling, Haihong Zhang, and Shuzhi Sam Ge. "Face recognition by applying wavelet subband representation and kernel associative memory." *IEEE Transactions on neural networks* 15, no. 1 (2004): 166-177.

[12] Yousefi, Bardia, Hina Hakim, Nasiha Motahir, Paria Yousefi, and Mohammad Mehdi Hosseini. "Visibility enhancement of digital dental X-ray for RCT application using Bayesian classifier and two times wavelet image fusion." *Journal of American Science* 8, no. 1 (2012): 7-13.

[13] Cherian, Sherimon Puliprathu, and Vinu Sherimon. "A proposed onto-Apriori algorithm to mine frequent patterns of high quality seafood." In *ICC*, pp. 169-1. 2017.

[14] Wu, Haotong. "Data association rules mining method based on improved apriori algorithm." In *Proceedings of the 4th International Conference on Big Data Research*, pp. 12-17. 2020.