RESEARCH ARTICLE   **OPEN ACCESS**

The Prediction Approach for Periodontitis Using Collaborative Filtering Method in Big Data

[1]SULAIGAL FEMINA.A, [2] R. SELVAMATHA M.Tech
[1]ME (CSE) FINAL YEAR
[2]ASSISTANT PROFESSOR

ABSTRACT
Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate. We have proposed 'the prediction approach' in big data which is usually based totally at the decision making or clinical acumen aspect of a Dentist. Collaborative filtering is a 'facts Mining method' in which data is fed into the machine which analysis it using certain parameters and offer resultant outcomes which helps in the future prediction of the disease, which in this case is Periodontitis. This will help in making decisions by examining or analyzing the whole disorder and plan the treatment to prevent its occurrence in future. It is by far the most specialized and latest technology used in Dentistry to prevent the disease occurrence in advance before it arise and create treatment strategies for future prevention of the disease. This method could be tested with the huge collection of ancient statistics of dental diseases to check the effectiveness of the technique and based on the generated end result; precision of prediction can be executed in future.
Keywords: Collaborative filtering, Big data analytics, periodontitis, Dental Anatomy.

I.INTRODUCTION

Periodontal diseases range from simple gum inflammation to serious disease that results in major damage to the soft tissue and bone that support the teeth. In the worst cases, teeth are lost. Whether your gum disease is stopped, slowed, or gets worse depends a great deal on how well you care for your teeth and gums every day, from this point forward. Periodontal (Gum) Disease What causes gum disease? Our mouths are full of bacteria. These bacteria, along with mucus and other particles, constantly form a sticky, colorless "plaque" on teeth. Brushing and flossing help get rid of plaque. Plaque that is not removed can harden and form "tartar" that brushing doesn't clean. Only a professional cleaning by a dentist or dental hygienist can remove tartar. Gingivitis The longer plaque and tartar are on teeth, the more harmful they become. The bacteria cause inflammation of the gums that is called "gingivitis." In gingivitis, the gums become red, swollen and can bleed easily. Gingivitis is a mild form of gum disease that can usually be reversed with daily brushing and flossing, and regular cleaning by a dentist or dental hygienist. This form of gum disease does not include any loss of bone and tissue that hold teeth in place. Periodontitis When gingivitis is not treated, it can advance to "periodontitis" (which means "inflammation around the tooth"). In periodontitis, gums pull away from the teeth and form spaces (called "pockets") that become infected. The body's immune system fights the bacteria as the plaque spreads and grows below the gum line.
Bacterial toxins and the body's natural response to infection start to break down the bone and connective tissue that hold teeth in place. If not treated, the bones, gums, and tissue that support the teeth are destroyed. The teeth may eventually become loose and have to be removed. Risk Factors • Smoking. Need another reason to quit smoking? Smoking is one of the most significant risk factors associated with the development of gum disease. Additionally, smoking can lower the chances for successful treatment. • Hormonal changes in girls/women. These changes can make gums more sensitive and make it easier for gingivitis to develop. • Diabetes. People with diabetes are at higher risk for developing infections, including gum disease. • Other illnesses. Diseases like cancer or AIDS and their treatments can also negatively affect the health of gums. • Medications. There are hundreds of prescription and over the counter medications that can reduce the flow of saliva, which has a

protective effect on the mouth. Without enough saliva, the mouth is vulnerable to infections such as gum disease. And some medicines can cause abnormal overgrowth of the gum tissue; this can make it difficult to keep teeth and gums clean. • Genetic susceptibility. Some people are more prone to severe gum disease than others. Who gets gum disease? People usually don't show signs of gum disease until they are in their 30s or 40s. Men are more likely to have gum disease than women. Although teenagers rarely develop periodontitis, they can develop gingivitis, the milder form of gum disease. Most commonly, gum disease develops when plaque is allowed to build up along and under the gum line.

How is gum disease treated?
The main goal of treatment is to control the infection. The number and types of treatment will vary, depending on the extent of the gum disease. Any type of treatment requires that the patient keep up good daily care at home. The doctor may also suggest changing certain behaviors, such as quitting smoking, as a way to improve treatment outcome.

Deep Cleaning (Scaling and Root Planing)
The dentist, periodontist, or dental hygienist removes the plaque through a deep-cleaning method called scaling and root planing. Scaling means scraping off the tartar from above and below the gum line. Root planing gets rid of rough spots on the tooth root where the germs gather, and helps remove bacteria that contribute to the disease. In some cases a laser may be used to remove plaque and tartar. This procedure can result in less bleeding, swelling, and discomfort compared to traditional deep cleaning methods.

Medications
Medications may be used with treatment that includes scaling and root planning, but they cannot always take the place of surgery. Depending on how far the disease has progressed, the dentist or periodontist may still suggest surgical treatment. Long-term studies are needed to find out if using medications reduces the need for surgery and whether they are effective over a long period of time. Listed on the next page are some medications that are currently used.
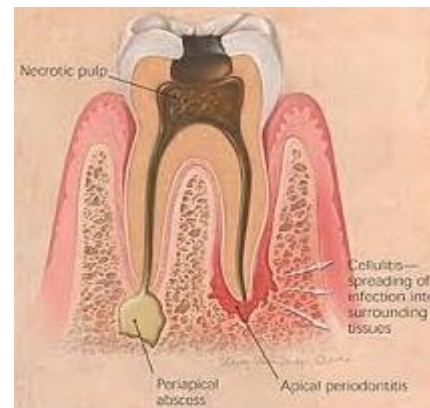
Who gets gum disease?
People usually don't show signs of gum disease until they are in their 30s or 40s. Men are more likely to have gum disease than women. Although teenagers rarely develop periodontitis, they can develop gingivitis, the milder form of gum disease. Most commonly, gum disease develops when plaque is allowed to build up along and under the gum line.
How do I know if I have gum disease?
Symptoms of gum disease include: • Bad breath that won't go away • Red or swollen gums • Tender or bleeding gums • Painful chewing • Loose teeth • Sensitive teeth • Receding gums or longer appearing teeth Any of these symptoms may be a sign of a serious problem, which should be checked by a dentist. At your dental visit the dentist or hygienist should: • Ask about your medical history to identify underlying conditions or risk factors (such as smoking) that may contribute to gum disease. • Examine your gums and note any signs of inflammation.

II. EXISTING SYSTEM
This is the first attempt to develop an alternative classification of periodontitis based on molecular profiles inperiodontitis-affected gingival tissues. It used an unsupervised machine learning approach to mine whole-genome gene expression data from a well-phenotype cohort of patients with either CP or AgP, and we identified two de novo clusters with highsimilarity in transcriptional profiles. Importantly, the clustersdid not align with the currently accepted periodontitis classification, but their distinct molecular signatures translated into discernible periodontitis-related characteristics: the two groupsdiffered significantly with respect to extent and severity of periodontitis, level of subgingival colonization, and intensity of serum antibody responses to periodontal micro biota. These data suggest that molecular profiling of gingival tissues can indeed form the basis for the development of an alternative, pathobiology based classification of periodontitis that correlates well with phenotypic features of the disease.

The study of tissue molecular profiles has led to significant insights in disease pathophysiology in several fields of medicine, most notably in oncology, where molecular profile-based classification of tumors showing different treatment response and prognosis has eventually translated into therapeutic decisions in clinical practice (Witt et al., 2011; Sadanandam et al., 2013).

A similar approach in the pathobiology and classification of human periodontitis has not been adapted to this date. It is important to emphasize that to extract sound, biologically meaningful data from transcriptional databases such as the one developed by our group, a clustering algorithm that can appropriately account for the underlying structure of the actual data must be utilized.

III. PROPOSED SYSTEM

We demonstrate a method for collaborative filtering of future Prediction. Previous work on recommender systems typically relies on feedback on a particular attribute of a disease, such as an age, and generalizes this to other items or other people. We examine the topic of unseen attribute recommendation through a user study of recommendation, where we aim to correctly estimate a filtering function for each user. Then by decomposing user parameters into shared and individual dimensions, we induce a similarity metric between users based on the degree to which they share these dimensions. We show that the collaborative filtering predictions of disease are more effective than pure content-based recommendation.

This is determined by the number of diseases which the patients have in common. In the most basic case, patients are removed only if they have no diseases in before each application of collaborative filtering, clustering is applied to the training set to discover connected components of patients. This served to remove the influence of patients who have little or no similarity with the testing patient for whom predictions are being common with the active patient. Thus, removing these patients does not result in loss of information, but effectively reduces the runtime of the algorithm. In practice, we cluster such that all patients in the training set have two or more diseases in common with the known diagnoses of the active patient. Introducing the constraint that clustering patients in the training set must have at least two common diseases with the active (testing set) patient enforces stronger similarities for all patients influencing the predictions. Essentially, we build a network of patients that are connected by at least two diseases and then perform collaborative filtering in this network. In theory, this helps to avoid the noise resulting from common diseases which introduce a very high number of weak influences. The clustering provides an additional benefit by reducing the number of diseases predicted on, which both simplifies and improves the collaborative filtering results.

(A) Algorithm 1: Item-based Filtering Enhanced by ESVD

Description We will now describe the steps of how ESVD can be combined with Item-based Filtering in order to make the base algorithm more scalable.

1. Define the original user-item matrix, R, of size $m \times n$, which includes the ratings of m users on n items. r_{ij} refers to the rating of user u_i on item i_j.

2. Preprocess user-item matrix R in order to eliminate all missing data values. The preprocessing is described in detail here:

(a) Compute the average of all rows, r_i, where i = 1, 2, ..., m, and the average of all columns, r_j, where j = 1, 2, ..., n, from the user-item matrix, R.

(b) Replace all matrix entries that have no values, denoted by $\perp$, with the corresponding column average, r_j, which leads to a new filled-in matrix, Rfilled−in.

(c) Subtract the corresponding row average, r_i, from all the slots of the new filled-in matrix, Rfilled−in, and obtain the normalized matrix Rnorm.

3. Compute the ESVD of Rnorm and obtain matrices U, S and V, of size $m \times m$, $m \times n$, and $n \times n$, respectively. Their relationship is expressed by: $R\,norm = U \cdot S \cdot V^T$.

4. Perform the dimensionality reduction step by keeping only k diagonal entries from matrix S to obtain a $k \times k$ matrix, S_k. Similarly, matrices U_k and V_k of size $m \times k$ and $k \times n$ are generated. The "reduced" user-item matrix, Rred, is obtained by $Rred = U_k \cdot S_k \cdot V^T_k$, while rr_{ij} denotes the rating by user u_i on item i_j as included in this reduced matrix.

5. Compute $\sqrt{S_k}$ and then calculate two matrix products: $U_k \cdot \sqrt{S_k}T$, which represents m users and $\sqrt{S_k} \cdot V^T_k$, which represents n items in the k dimensional feature space. We are particularly interested in the latter matrix, of size $k \times n$, whose entries represent the "meta" ratings provided by the k pseudo-users on the n items. A "meta" rating assigned by pseudo-user u_i on item i_j is denoted by mr_{ij}.

6. Proceed with Neighborhood Formation which can be broken into two substeps: (a) Calculate the similarity between items i_j and i_f by computing their Adjusted Cosine Similarity as follows:

$$sim_{jf} = adjcorr_{jf} = \frac{\sum_{i=1}^{k} mr_{ij} \cdot mr_{if}}{\sqrt{\sum_{i=1}^{k} mr_{ij}^2 \sum_{i=1}^{k} mr_{if}^2}}$$

where k is the number of pseudo-users, selected when performing the dimensionality reduction step. We have to note a change between the Adjusted Cosine Similarity equation utilized in plain Item-based Filtering and here. In plain Item-based Filtering the difference in rating scale between distinct users was offset by subtracting the corresponding user average from each co-rated pair of items. In ESVD-enhanced Item-based Filtering, that difference in rating scale was offset during the normalization of the original user-item matrix which yielded matrix R_{norm}.

(b) Based on the results from the Adjusted Cosine Similarity calculations for pairs of items including the active item and a random item, isolate the set of items which appear to be the most similar to the active item.

7. Conclude with Prediction Generation, achieved by the following weighted sum:

$$pr_{aj} = \frac{\sum_{k=1}^{l} sim_{jk} * (rr_{ak} + \bar{r_a})}{\sum_{k=1}^{l} |sim_{jk}|}$$

which calculates the prediction for user u_a on item i_j. It is similar to the equation utilized by plain Item-based Filtering in that it bases its predictions on the ratings given by the active user, u_a, on the l items selected as the most similar to active item i_j . Yet, it is different in that the user ratings are taken from the reduced user-item matrix, R_{red}. Also, we have to add the original user average, $\bar{r_a}$, back since it was subtracted during the normalization step of the preprocessing.

(B) FLOW OF MODULES
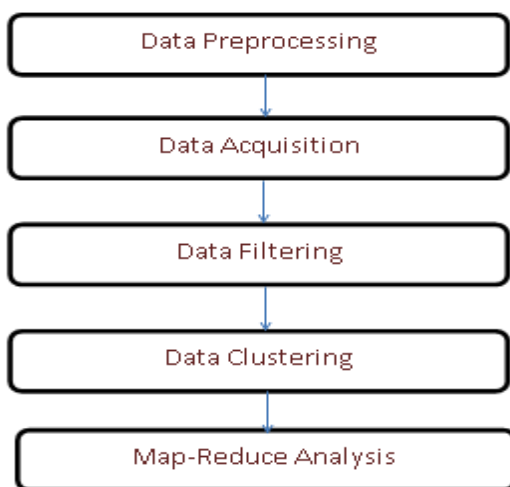
Data Acquisition
Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes. In this, data acquisition process dataset including periodontitis affected value. Here, periodontitis affected datasets have been collected for processing them on big data.

Data Filtering
Data filtering will involve taking out information that is useless to a reader or information that can be confusing. Generated reports and query results from dataset often result in large and complex data sets. Redundant or impartial pieces of data can confuse or disorient a user. Filtering data can also make results more efficient. The data acquisition process where collected dataset been viewed by user. The viewed data will be filtered for clustering.

Clustering
Clustering divides the data into groups that are meaningful. If meaningful groups are the goal, then the cluster should capture natural structure of the data. The proposed solution uses Item Based collaborative filtering for clustering the data. Thereby the various levels of clustering will be processed.

Map Reduce
Map Reduce is a framework for processing parallelizable problems across huge datasets using a large number of computers (nodes), collectively referred to as a cluster a grid

- **Map step:** Each worker node applies to the local data, and writes the output to a temporary storage. A master node ensures that only one copy of redundant input data is processed.
- **Shuffle step:** Worker nodes redistribute data based on the output keys, such that all data belonging to one key is located on the same worker node.
- **Reduce step:** Worker nodes now process each group of output data, per key, in parallel.

Statistical Report
A statistical report is the gathering of data and the generation of conclusions from a periodontist. This could be people's opinions and thoughts on a product or items that result in a survey. Here the statistical report will be generated in R programming. The separate statistical report will be generated for Periodontitis affected value, partially affected periodontitis value and for normal person.

IV. CONCLUSION AND FUTURE ENHANCEMENT
Finally, a Clubcf (Collaborative Filtering) algorithm is applied. The strategy will filter most of the unwanted knowledge with the formula in the

initial section. Then using the Clubcf method the massive knowledge is analysed using specific algorithms. The collaborative filtering is a statistics Mining technique, which has been derived to approaches the input fed to the system which in turn provides resultant results to expect the future track or path of the sickness.

This work has the distinctiveness in that it studies or analyses the complete disease and helps prevention by forecasting its future trend and it's specialized to save you from practice errors in Dental diseases prediction, in advance before it occur and make disease prevention a reality. This work may be tested with the collective information of dental database archives to check the effectiveness of the system and based totally on the generated result; future prediction may be carried out to strategize treatment plans.

REFERENCES

[1] Armitage GC (2010). "Comparison of the microbiological features of chronic and aggressive periodontitis," Periodontol, 2000, 53:70-88.

[2] Benjamini Y, Hochberg Y. 'Controlling the false discovery rate: a practical and powerful approach to multiple testing," 1995, 57:289-300.

[3] Chaushu S, Wilensky A, Gur C, Shapira L, Elboim M, Halftek G, et al. "Direct recognition of Fusobacteriumnucleatum by the NK cell natural cytotoxicity receptor NKp46 aggravates periodontal disease," PLoSPathog , 2012, 8:e1002601.

[4] Claudio N, Dalet A, Gatti E, Pierre P , "Mapping the crossroads of immune activation and cellular stress response pathways." 2013, 32:1214-1224.

[5] Fraley C, Raftery AE, Murphy TB, Scrucca L :" normal mixture modeling for model-based clustering, classification, and density estimation. Technical report 597. Seattle, WA: University of Washington," Department of Statistics, 2012, MCLUST version 4 .

[6] Hernandez M, Dutzan N, García-Sesnich J, Abusleme L, Dezerega A, Silva N, et al." Host-pathogen interactions in progressive chronic periodontitis," 2011, 90:1164-1170.

[7] O. Sporns and M. P. van den Heuvel, "Network maps of the human brain's rich club," Network Science, 2013, vol. 1, no. 2, pp. 248-250.

[8] SrinivasLaxminarayan, Mark J. Buller, William J. Tharion, and JaquesReifman, "Human Core Temperature Prediction for Heat-Injury Prevention," IEEE Journal Of Biomedical And Health Informatics, May 2015, Vol. 19, No. 3.