Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 215 (2022) 519–528

Procedia Computer Science

www.elsevier.com/locate/procedia

4th International Conference on Innovative Data Communication Technology and Application

A Comparative Study of Machine Learning Regression Approach on Dental Caries Detection

Chandan Kumar[a*], Bijay Singh[b]

[a]Amrita School of Computing, Amrita Vishwa Vidyapeetham, Amaravati Campus, Amaravati-522203, India
[b]School of Engineering & IT, Arka Jain University, Jamshedpur-832108, India

Abstract

Cavities are the most prevalent consequence of dental caries, an infectious condition that weakens the structure of the teeth and may be spread from person to person. Research on dental caries, which is considered one of the most widespread problems with oral health, has been conducted with the purpose of early diagnosis owing to the discomfort and expense of treatment. In recent years, artificial intelligence has been utilized to construct models for estimating the probability of dental caries. In the current research work, the Machine learning technique has been provided as the prominent solution in providing the prediction model to detect dental caries. This research work reports a few machine learning algorithms such as Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), and Naïve Bayes (NB) for providing a model for dental caries detection. Finally, the said algorithms are evaluated over the influencing parameters such as accuracy, precision, recall, F1-Score, and Mathews Correlation Coefficient (MCC). The empirical analysis shows that the DT provides a more accurate model with an accuracy level of 85.62%.

Keywords: Dental Caries; ML; RF; DT; LR; NB

1. Introduction

Caries in the teeth is a common infectious oral illness that may last for years and affects both adults and adolescents all over the globe. The fact that there are roughly 1.8 billion occurrences of illness linked to caries reported each year suggests that a decline in dental health may eventually lead to greater records of cases of tooth loss [1]. Caries is seen in a much higher percentage of children and teenagers as they become older. It demonstrates an epidemiologic feature

10.1016/j.procs.2022.12.054

that gradually persists into adolescence, after which tooth decay increases [2]. Untreated dental caries at the very beginning i.e., from childhood and, it is imperative that dental caries be treated from the infancy stage. Therefore, since it has an effect on the quality of life as a whole, it is regarded as an issue that impacts the public's overall health rather than an individual's health [3]. If dental cavities are detected in their earliest stages, patients have the opportunity to receive both preventative and restorative treatment options. Long-term procedures such as root canal therapy, dental prosthetics, and full-mouth reconstruction can be necessary for cleaning if one ignores his/her teeth for a lengthy period of time and damage the dentin or pulp [4]. It is highly recommended that everyone gives their dental health the care it deserves and schedules frequent appointments with their dentist for examination and medical care. However, due to financial limitations or a lack of awareness of the importance of oral health, prevention or evaluation is frequently neglected before the beginning of symptoms. [5]. This might be a problem since oral health is so important. The number of permanent teeth that have been impacted by caries is tallied using the decayed–missing–filled teeth index. This index takes into account both teeth that are missing and teeth that have decayed. This index is the leading population-based assessment of ailments linked to dental caries globally (missing or decayed). Inappropriate oral health management behaviors, dietary habits, dental care utilization, and economic factors may all harm oral health [6]. Therefore, determining DMF (decaying, missing, filled) teeth and the variables that are connected to it, as well as making a prediction about whether or not a person will develop dental caries in their permanent teeth, may serve as an important foundation for developing an individual oral preventive plan. The goal of the Oral Health Act assessment of children's oral health is to gather the information that may be used to better control dental caries in children. The study investigates the condition of oral health in the general population, in addition to accompanying paperwork and the use of dental care goods. The survey results in the area of oral diseases that was conducted in 2018 surveyed a total of 27,568 participants, which is the benchmark for what is known as "big data." According to the results of the research, DMFT has a propensity, with increasing age to noticeably increase.

Machine learning (ML) has evolved into an indispensable tool for comprehending and analyzing data as extensive as the survey described earlier, and it is now being used in the realm of medicine in a variety of various ways. ML provides a way for future prediction through learning. Making an ML-based model requires two different phases such as the training and testing phase. The learning is performed in the training phase and based on the training the prediction is made in the testing phase [7]. ML's beginnings can be seen in the 1950s, continued to advance through the 1980s and 1990s, and then came to a standstill after that. By the mid-2000s, the Internet had amassed a large amount of big data, and equipment had been built to handle it [8]. In recent years, great advancements have been achieved in the use of machine learning for image identification, voice recognition, and translation. A study on deep learning (DL) is now being carried out in the area of dentistry in Korea [9-10]. This research makes use of the dental caries data consideration for applying the machine learning approaches.

A. *Motivation*

The conventional way to predict dental caries is to find the DMFT factor. But the main limitation present behind the conventional solution is computational time. In contrast to the traditional way, prior to an expert performing a test, (ML) can predict carious teeth in a population using surveys or other basic data for full diagnostic on the patient. The amount of time, money and human resources necessary for an evaluation of dental health will be cut down significantly. In addition to this, ML may assist in the classification of high-risk patients so that they can be provided with an appropriate diagnosis and the required treatment by an expert. Furthermore, if the predictive model can pinpoint important factors that have a considerable influence on the development of caries, then the prevention of caries may be achieved by the management of those factors.

B. *Objective*

The objectives of this paper are summarized as follows:

- To explore machine learning techniques.
- To implement state-of-the-art machine learning algorithms to detect dental caries.
- To measure the performance level with respect to some influencing parameters such as accuracy, precision, recall, and F-1 score.

C. *Paper structure*

The rest of this paper is structured in the following way. Section 2 reports the literature survey. Section 3 shows the materials and methods used in the research work. Section 4 presents the result and discussion. Finally, the conclusion is reported in Section 5.

Nomenclature

AI	Artificial Intelligence
CAD	Computer-Aided Diagnosis
CAL	Clinical Attachment Loss
CNN	Convolutional Neural Network
DL	Deep Learning
DT	Decision Tree
DMFT	Decaying, Missing, Filled, Teeth
EDA	Explorative Data Analysis
LR	Logistic Regression
MCC	Mathews Correlation Coefficient
ML	Machine Learning
NB	Naive Bayes
RF	Random Forest

2. Literature Survey

In the healthcare system, the use cases of Artificial Intelligence (AI) include a patient monitoring system, collecting the data from the patient through sensors, and a decision support system for handling big data related to health information. AI can play a vital role in proving some automotive models to diagnose the disease at its early stage. ML may help physicians identify and treat patients by evaluating real-time data to discover changes.

2.1. Deep Learning Approach

According to the survey, the diagnosing of disease through medical images is at the highest ML-based healthcare system growth rate. A significant amount of information on medical pictures is also included which is easily available, which makes it simpler to do research on this topic. Many AI solutions have been established in the medical field based on this paradigm to ease the time-consuming and cost-related aspects of conducting certain diagnoses. By combining and evaluating data from imaging tests like X-rays and CT, VUNO, an example case, assists in bone age evaluation using DL. Lunit builds an AI-based solution for chest X-rays and mammography images using DL technology that enhances early detection and lowers false diagnosis rates [11].

The advantage of current methods based on DL, a subset of ML, is learning to automatically extract crucial properties from images. [12]. The most widely used approach for segmenting, classifying, and detecting organs or disorders in medical pictures is the convolutional neural network (CNN) [13]. Dental and maxillofacial surgeons have used CAD and CNN to identify landmarks in cephalograms, identify teeth, classify teeth, diagnose caries, and detect maxillary sinusitis. [14]. By incorporating a second opinion into the information obtained through this process, a dentist may provide a more precise and fast diagnosis. Recently, research on RBL recognition using CNN in dental panoramic images has also been conducted. [15].

2.2. Machine Learning Approach

Oral health has emerged as one of the most crucial factors in determining the overall quality of one's life in today's rapidly aging society. The sixth most prevalent disease worldwide is periodontitis, and it is one of the illnesses that are most prevalent among human beings. It is an infection of the gums that surrounds the teeth. One of the most obvious signs of periodontitis is the loss of alveolar bone [16]. The loss of bone eventually leads to tooth loss, an edentulous jaw, and issues with food chewing. To study the origin and treatment of periodontitis, it was important to create a classification system for the condition. [17]. Since then, periodic updates to the periodontitis categorization system have been made to reflect the most recent developments in both clinical and scientific knowledge [18]. A periodontal probe-based assessment of periodontal health called Clinical Attachment Loss (CAL) is used. Because this approach is unreliable, a computer-aided diagnosis (CAD) system may assist physicians in making judgments by identifying significant information from medical photographs acquired in a variety of settings [19]. However, owing to the variety of illness patterns, feature extraction is difficult and time-consuming in the conventional CAD technique [20].

The contrast of the X-ray image is the major variable that is controlled in studies on the classification of dental caries based on image interpretation. The authors of [21] describe a unique method for detecting the presence of dental caries and cysts in dental images using a hybridized negative transformation and statistical analysis. Cavities were analyzed using the radon transform and discrete cosine transform, while other research used classification approaches like decision trees, random forests, and naive Bayes to diagnose. By altering the picture contrast, this approach enhanced the diagnostic accuracy of caries, however, the study's drawback was highlighted owing to the formation of image noise. To address picture noise restrictions, the histogram equalization approach was used [22-23]. To facilitate early caries identification in clinics, the authors of [24] created a kernel-modified SVM and a watershed image enhancement method.

2.3. Critical Finding

The existing literature shows that for dental caries prediction two approaches are preferably chosen such as the deep learning approach and the machine learning approach. For the deep learning approach, the primary input is the image data such as X-RAY, CT scan reports, etc. But machine learning approaches are mostly used numeric data values.

3. Materials and Methods

In this paper, we applied only machine learning regression approaches. The dataset and the methods employed in the current study endeavor have been detailed in this part.

3.1. Dataset

The dental dataset considered during the research work has been taken from the UCI machine learning data repository. It is having 16 features and 7456 samples with the class “EXAM SVC ANNOTATION CODE” having values as normal and treatment. Table 1 shows the dataset description. Figure 1 shows the class count of the dataset.

Table 1. Dataset description.

Feature	Sample	Class
16	7456	Normal= 6098, Treatment= 1358

3.2. Methodology

Figure 2 shows the workflow of the methodology. Initially, the dataset has been considered for the data preprocessing. To the preprocessed data, Explorative Data Analysis (EDA) has been done in order to visualize the dataset. The dataset is split into the training and testing sets with a distribution factor of 0.3. 70% of the dataset is considered for the training set and the remaining 30% is considered for the testing phase. Out of 7456 samples, 5219 samples have been considered from the training phase from which the class count for normal, and treatment are 4306 and 913 respectively. For testing phase 2237 samples are considered in which the class count for normal, and treatment are 1792 and 445 respectively. To the preprocessed dataset state-of-the-art machine learning algorithms such as Random Forest, Decision Tree, Naïve Bayes, and Logistic Regression techniques are applied, and the performance analysis has been done in order to find out the best-fit algorithm.

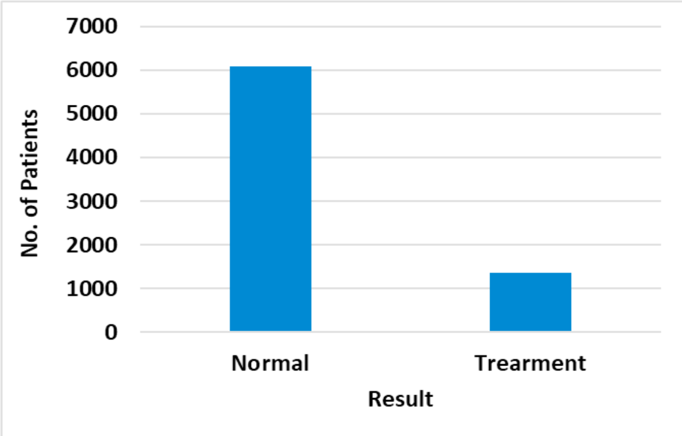

Fig. 1. Class count for the dataset

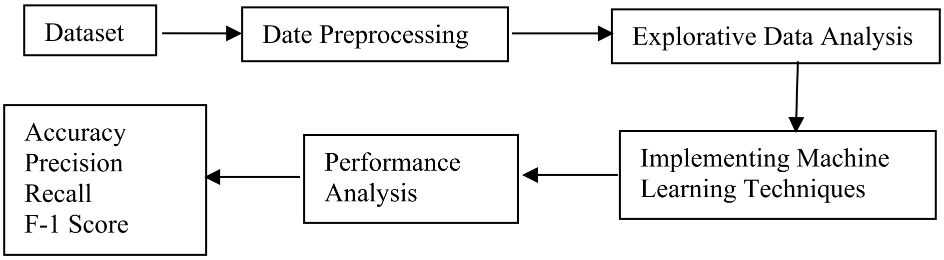

Fig. 2. Workflow of the adopted methodology

1) *Random Forest:*

Random Forest (RF) is a well-known ensemble technique in the machine learning concept. It comprises of several Decision Trees. The randomization of the selection criteria used in the node split is what ultimately leads to the achievement of diversity. In this method, any random characteristic is selected for the dividing purpose rather than selecting the most advantageous feature. It is a strategy known as bagging, in which many deep trees are joined together to provide an output with a low variation. The RF will run the Decision Tree m times when given an input A, where A is a vector made up of many qualities, and then take the average of the results to determine the actual prediction.

2) *Decision Tree*

The decision tree, which is given a collection of qualities, attempts to generate a rule based on the categorization that may be carried out. In a decision tree, the procedure will typically begin with a root, and then progressively progress to the point where the data characteristics are split according to the information gain score. The DT will have three different kinds of nodes: the root node, the test node, and the leave node, which is also called the decision node. The construction of a DT involves the use of a classifier (ID3, CART, CHAID, etc.) that, in general, constructs a tree automatically. The primary goal of the classification method, however, is to produce a decision tree that is optimally suited to the data that has been provided. It does this by using a variety of methods in order to come to a conclusion on whether or not a single node may be split into two or more sub-nodes. The fundamental benefit of DT is that it needs very little data preparation before it can be used, and it is also capable of handling both numerical and categorical data. This is a significant advantage over other methods. However, the most significant drawback of using this method of categorization is overfitting.

3) *Naïve Bayes*

The Naïve Bayes (NB) classification method derives its functionality from the conditional probability theorem developed by Bayes. Within the context of this discussion, probability refers to the level of belief. The prediction of the class c from feature vector F {f1,f2,f3,…fn}. The conditional probability of c can be defined by equation 1.

$$P(c|f1, f2, .., fn) = \frac{\{P(f1|c) * P(f2|c) *..P(fn|c)\} * P(c)}{P(f1) * P(f1) *..P(fn)} \quad (1)$$

The data are being classified by the use of the conditional probability. The fact that this algorithm operates on the premise that each of the traits is unrelated to the others is perhaps the aspect of it that is of the greatest significance. There are three distinct varieties of algorithms that are based on NB, and these are the Bernoulli NB, the Gaussian NB, and the Multinominal NB. The most significant benefit of this classification is that it only needs a relatively little amount of training data in order to estimate the conditional parameters; nevertheless, the amount of time needed for this estimation is contingent on the size of the dataset. If the dimension of the dataset is really large, then the NB will perform poorly as an estimator since the amount of time and cost required to estimate will rise proportionally with the size of the dataset.

4) *Logistic Regression*

The regression model is what may be utilized to make predictions about the likelihood of the data that has been presented. The logistic function, also called the sigmoid function, is a well-defined model that is essential to its operation. Its operation is dependent on this model. The probabilities in this model define the possible outcomes of a single trial, and the sigmoid function is used to tune them. The sigmoid function converts any real number to a probability value between one and zero. The primary benefit of this classification model is that it can be easily implemented on independent variables; the challenge lies in determining which variable is independent when working with a high-dimensional data set [25].

4. Result and discussion

This section reports the environmental setup used for the implementing methodology. In addition, this section shows the EDA and performance analysis of the research work.

4.1. Environment Definition.

For implementing RF, DT, LR, and NB the system has 500 GB SSD, 1 TB HDD, 8 GB RAM, Intel i3 processor having a clock speed of 2.4-3.0 GHz and Weka 3.8 have been considered.

4.2. Explorative Data Analysis:

Figures 3, 4, and 5 are the EDA plots of the considered dataset. Figure 3 shows the correlation plot among different features present in the dataset. A correlation heatmap is a graphical depiction of a correlation matrix that depicts the degree to which several variables are linked to one another. Any number between -1 and 1 may be assigned to the value of correlation. A causal connection does not always follow from the existence of a correlation between two random variables or between bivariate data. Figure 4 shows the number of values for different attributes with respect to the class variable. Figure 5 shows the scatter plot among different features of the dataset.

4.3. Performance Analysis

The influencing parameters of the current study include the accuracy, precision, recall, F-1 score, and Mathew's co-relation coefficient as defined in equations 2-6.

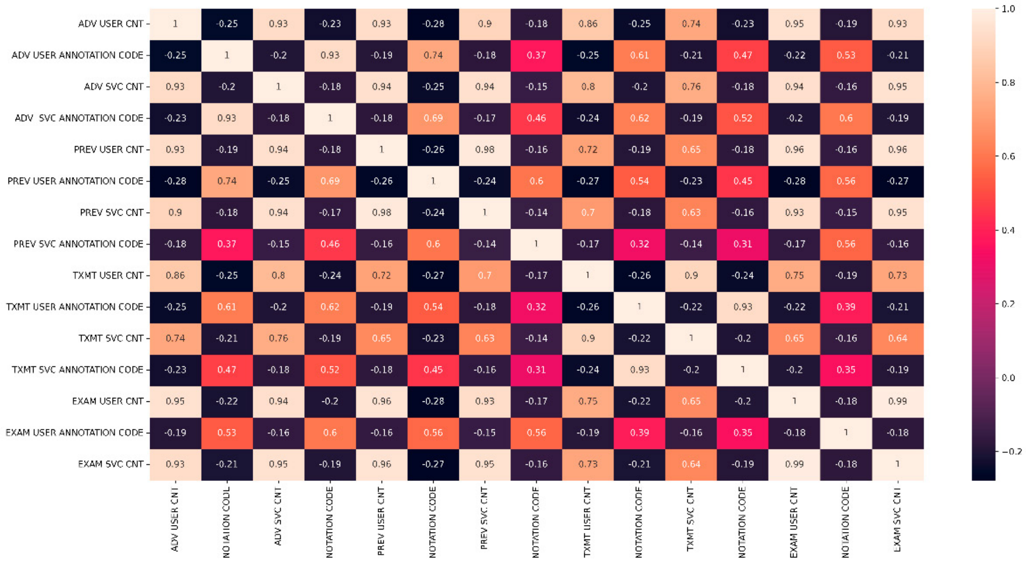

Fig. 3. Correlation heatmap

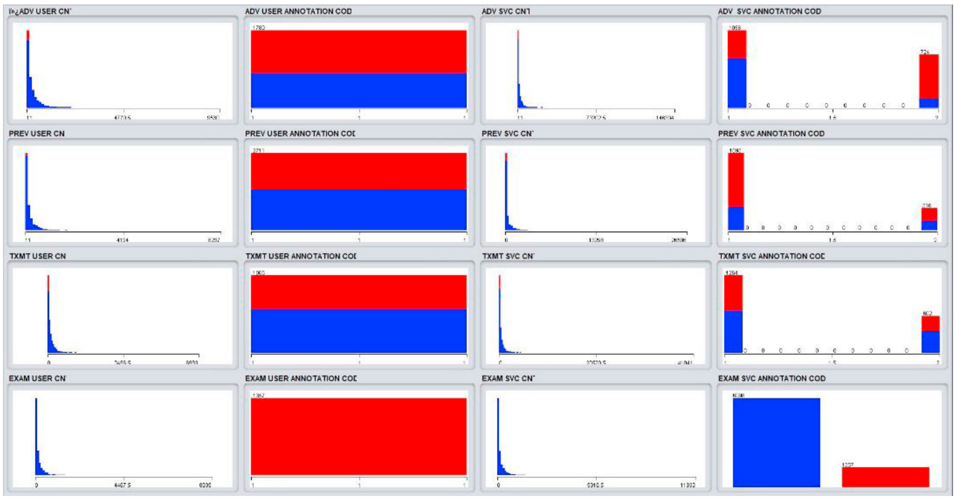

Fig. 4. Feature value count with respect to class variable

Accuracy: Accuracy can be defined as the ratio of the sum of the true positive and negative values to all predicted values. Equation (2) provides the measurement that establishes a model's effectiveness at recognizing trends and connections between characteristics in the datasets, where TN (true negative) refers to a negative observation that was successfully anticipated to be negative and is calculated as the ratio of the correctly predicted value to the total number of assessments.

$$Accuracy = \frac{TR_a + TR_b}{TR_a + TR_b + FL_a + FL_b} \qquad (2)$$

Precision: It is used to measure the correctly predicted values which is shown in equation (3). Equation (3) states that a classifier's capacity to recognize the pertinent data points. When a positive observation is also expected to be positive, this is known as a TP (true positive). False positives are negative observations that are reported as positives, or FPs.

$$Precision = \frac{TR_a}{TR_a + FL_a} \quad (3)$$

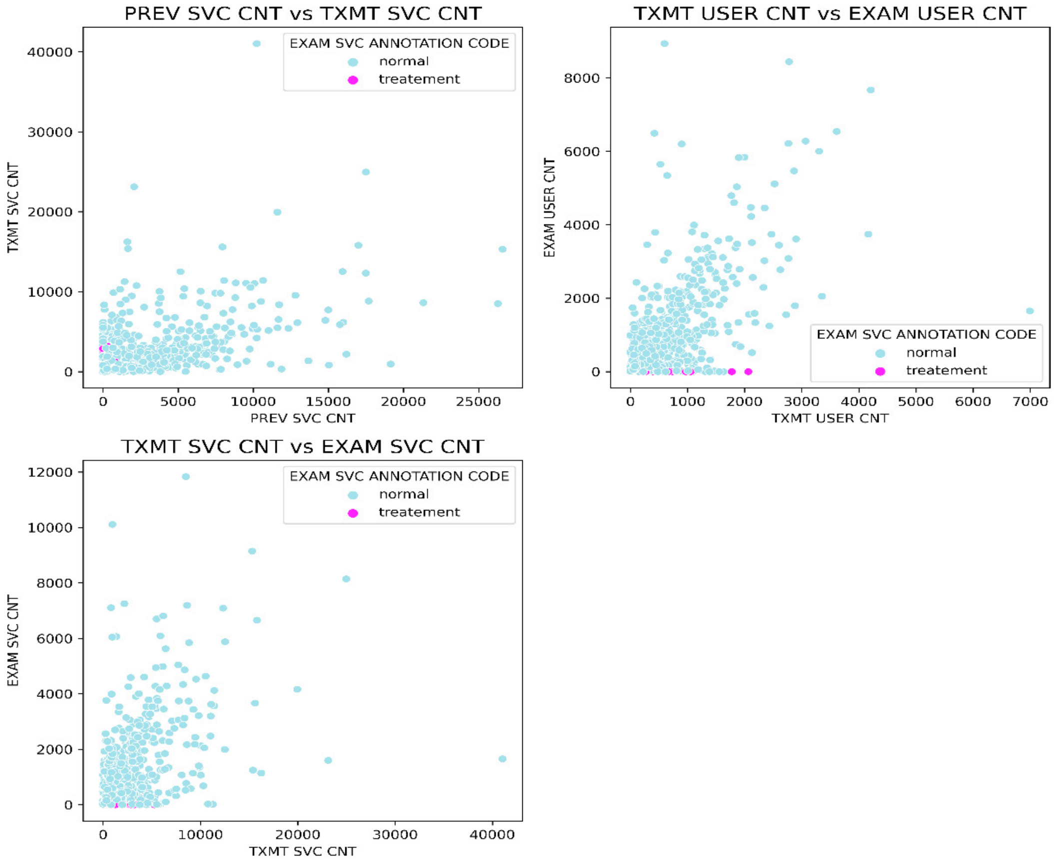

Fig. 5. Scatterplot of different attributes

Recall: It is calculated as the true predicted values out of actual predicted positive value. Equation (4) provides the measurement for determining how accurately a model has discovered the effective classes in our sample. False negative, or FN, is an observation that was favourable but was expected to be negative.

$$Recall = \frac{TR_a}{TR_a + FL_b} \quad (4)$$

F1-Score: It is the harmonic mean of precision and recall value The accuracy of the model on the dataset is gauged by the F1-Score, which is shown in Equation (5).

$$F1 - Score = 2\,x\,\frac{\left(\frac{TR_a}{TR_a + FL_a} * \frac{TR_a}{TR_a + FL_b}\right)}{\frac{TR_a}{TR_a + FL_a} + \frac{TR_a}{TR_a + FL_b}} \quad (5)$$

MCC: It is a statistical tool for evaluating models. Its responsibility is to evaluate or quantify the discrepancy between the projected and actual values. The calculation process of MCC is shown in equation (6).

$$MCC = \frac{(TR_a * TR_b) - (FL_a * FL_b)}{\sqrt{(TR_a + FL_a)} + (TR_a + FL_b) + (TR_b + FL_a) + (TR_b + FL_b)} \quad (6)$$

The TR_a, TR_b, FL_a, and FL_b are the true positive, true negative, False Positive and False Negative respectively. Equation 2-6 shows the definition of the performance parameter. Table 2 shows the performance parameter of the adopted RF, DT, LR and NB. Table 3 shows the computational time for each model. Figure 6 shows the comparison of performance measure of the implemented algorithms.

Table 2. Performance Evaluation

Method	Accuracy	Precision	Recall	F-1 Score	MCC
RF	84.09	84.14	97.89	90.5	48.5
DT	85.62	84.86	97.67	90.82	61.5
NB	77.93	80.54	90.44	85.20	43.4
LR	79.02	82.60	89.43	85.88	45.8

Table 3. Computational Time for each model

Method	Computational Time (sec)
RF	76
DT	65
NB	89
LR	112

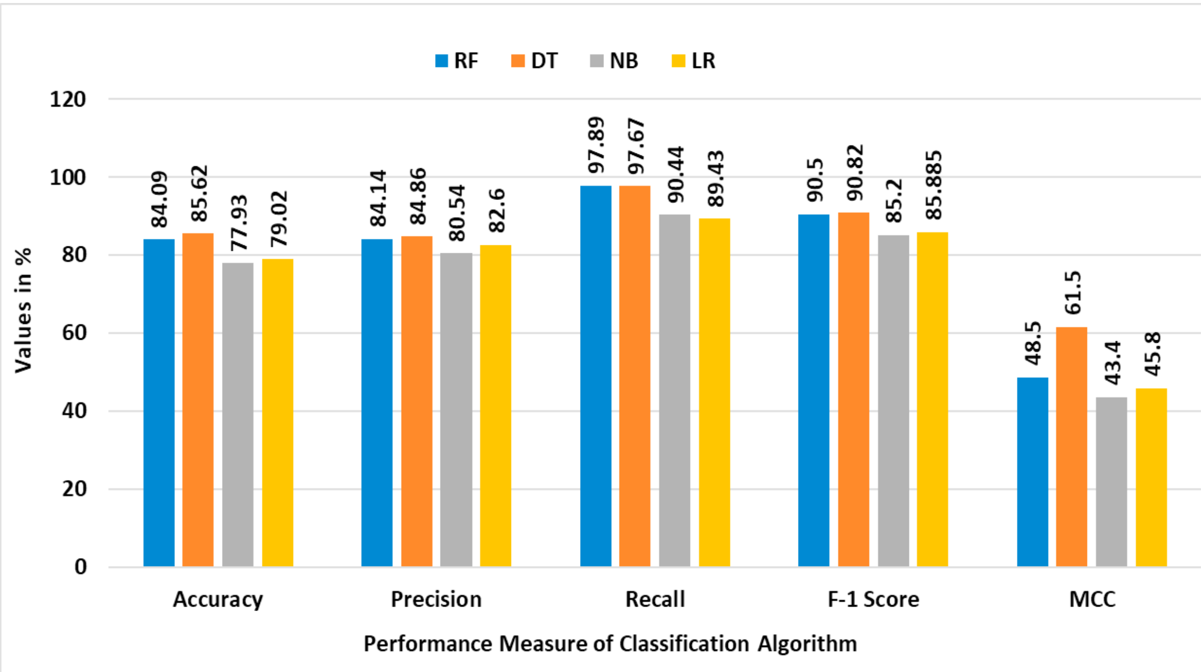

Fig. 6. Comparison of Performance Measures

5. Conclusion

In this research work, various machine learning algorithms such as RF, DT, LR and NB have been implemented on a public dental data repository for dental caries detection. For measuring the performance level of every algorithm, the parameters such as accuracy, precision, recall, F1- Score, and MCC have been considered. The empirical analysis done during the research work shows that the DT shows highest performance with accuracy as 85.62, In addition to accuracy the other parameters are precision, recall, F1-Score and MCC are 84.86%, 97.67%, 90.82%, and 61.5% respectively.

References

[1] Dye, B.A.; Li, X.; Thorton-Evans, G. Oral Health Disparities as Determined by Selected Healthy People 2020 Oral Health Objectives for the United States, 2009–2010; National Center for Health Statistics: Washington, DC, USA, 2012; Volume 104, pp. 1–8.

[2] Korea National Children's Oral Health Survey. 2018. Available online: https://www.korea.kr/common/download.do?tblKey=EDN&fileId=188769457

[3] Cummins, D. Dental caries: A disease which remains a public health concern in the 21st century—The exploration of a breakthrough technology for caries prevention. J. Clin. Dent. 2013, 24, 1–14.

[4] Aggeryd, T. Goals for oral health in the year 2000: Cooperation between WHO, FDI and the national dental associations. Int. Dent. J. 1983, 33, 55–59.

[5] Kim, A.H.; Han, S.Y.; Kim, B.I.; Kim, H.D.; Kwon, H.K. The characteristics of high caries risk group for 12-year-old children in Korea. Korean Acad. Prev. Dent. Oral Health 2010, 34, 302–309.

[6] Hwang, D.-H.; Lee, J.-H. A study on DMFT index between elementary school students: Korea national health and nutrion examination survey. J. Korean Soc. Oral Health Sci. 2021, 9, 1–6.

[7] Seul, M.S. Current status and future developments of machine learning artificial intelligence in law: Focusing the cusp of machine learning in U.S. and discourses over legal profession and law school education. Justice 2016, 156, 269–302.

[8] Velarde, G. Artificial Intelligence and its Impact on the Fourth Industrial Revolution: A Review. Int. J. Artif. Intell. Appl. 2019, 10, 41–48.

[9] Lee, J.-H.; Kim, D.-H.; Jeong, S.-N.; Choi, S.-H. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. J. Dent. 2018, 77, 106–111.

[10] Lee, D. Measurement and Evaluation of Bone Density for Dental Implant Surgery Using Synthetic Ghost-free Panoramic Radiograph; Seoul National University: Seoul, Korea, 2020.

[11] Lunit. "Lunit INSIGHT CXR1, Lunit INSIGHT CXR2, Lunit INSIGHT MMG". Available online: https://insight.lunit.io/ (accessed on 26 December 2021).

[12] Kallenberg, M., Petersen, K., Nielsen, M., Ng, A.Y., Diao, P., Igel, C., Vachon, C.M., Holland, K., Winkel, R.R., Karssemeijer, N. and Lillholm, M., 2016. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. IEEE transactions on medical imaging, 35(5), pp.1322-1331.

[13] Hannun, A.Y.; Rajpurkar, P.; Haghpanahi, M.; Tison, G.H.; Bourn, C.; Turakhia, M.P.; Ng, A.Y. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat. Med. 2019, 25, 65–69.

[14] Kim, J.; Lee, H.-S.; Song, I.-S.; Jung, K.-H. DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs. Sci. Rep. 2019, 9, 17615.

[15] Chen, C.; Bai, W.; Davies, R.H.; Bhuva, A.N.; Manisty, C.H.; Augusto, J.; Moon, J.; Aung, N.; Lee, A.M.; Sanghvi, M.M.; et al. Improving the Generalizability of Convolutional Neural Network-Based Segmentation on CMR Images. Front. Cardiovasc. Med. 2020, 7, 105.

[16] Tonetti, M.S.; Jepsen, S.; Jin, L.; Otomo-Corgel, J. Impact of the global burden of periodontal diseases on health, nutrition and wellbeing of mankind: A call for global action. J. Clin. Periodontol. 2017, 44, 456–462.

[17] Armitage, G.C. Development of a Classification System for Periodontal Diseases and Conditions. Ann. Periodontol. 1999, 4, 1–6.

[18] Caton, J.G.; Armitage, G.; Berglundh, T.; Chapple, I.L.; Jepsen, S.; Kornman, K.S.; Mealey, B.L.; Papapanou, P.N.; Sanz, M.; Tonetti, M.S.; et al. A new classification scheme for periodontal and peri-implant diseases and conditions—Introduction and key changes from the 1999 classification. J. Periodontol. 2018, 89, S1–S8.

[19] Kwon, O.; Yong, T.-H.; Kang, S.-R.; Kim, J.-E.; Huh, K.-H.; Heo, M.-S.; Lee, S.-S.; Choi, S.-C.; Yi, E.-J. Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep con-volution neural network. Dentomaxillofac. Radiol. 2020, 49, 20200185.

[20] Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: An overview and application in radiology. Insights Imaging 2018, 9, 611–629.

[21] Veena, D.K.; Jatti, A.; Joshi, R.; Deepu, K.S. Characterization of dental pathologies using digital panoramic X-ray images based on texture analysis. In Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Korea, 11–15 July 2017; pp. 592–595.

[22] ngh, P.; Sehgal, P. Automated caries detection based on Radon transformation and DCT. In Proceedings of the 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, India, 3–5 July 2017; pp. 1–6.

[23] Ghaedi, L.; Gottlieb, R.; Sarrett, D.C.; Ismail, A.; Belle, A.; Najarian, K.; Hargraves, R.H. An automated dental caries detection and scoring system for optical images of tooth occlusal surface. In Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 1925–1928.

[24] Kuang, W.; Ye, W. A Kernel-Modified SVM Based Computer-Aided Diagnosis System in Initial Caries. In Proceedings of the Second International Symposium on Intelligent Information Technology Application, IEEE, Shanghai, China, 20–22 December 2008; Volume 3, pp. 207–211.

[25] Andi, H. K. . An Accurate Bitcoin Price Prediction using logistic regression with LSTM Machine Learning model. Journal of Soft Computing Paradigm, 2021, 3(3), 205-217.