



The precision of case difficulty and referral decisions: an innovative automated approach

Shivani Mallishery¹ · Pavan Chhatpar² · K. S. Banga³ · Trusha Shah³ · Pankaj Gupta³

Received: 1 April 2019 / Accepted: 5 August 2019 / Published online: 13 August 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Objectives Endodontic treatment works as a successful treatment modality in several cases. However, it may fail due to some reasons unforeseeable by the dentist. Many failures can be prevented by carefully assessing the difficulty level of the case before initiating treatment or by referral to a specialist. This study presents an approach using machine learning to generate an algorithm which can help predict the difficulty level of the case and decide about a referral, with the help of the standard American Association of Endodontists (AAE) Endodontic Case Difficulty Assessment Form.

Materials and methods Using the AAE Endodontic Case Difficulty Form after obtaining the patients' consent, 500 potential root canal patients were diagnosed. The filled forms were assessed by two pre-calibrated endodontists, and, in cases of conflicting opinion, a third endodontist's opinion was taken. Artificial neural network was used for generating the algorithm.

Results Using 500 filled AAE forms, a sensitivity of 94.96% was achieved by the machine learning algorithm.

Conclusion This study provides an option for automation to the conventional method of predicting the difficulty level of a case, thus increasing the speed of decision-making and referrals if necessary.

Clinical relevance An AAE Endodontic Case Difficulty Assessment Form when utilized along with machine learning can assist general dentists in rapid assessment of the case difficulty. This is a helpful tool in developing countries, where endodontic treatment and referral guidelines are often neglected. It also helps to make difficulty level assessments easier for novice practitioners, when they are in doubt about the same.

Keywords Artificial intelligence · Case difficulty · Machine learning · Referral · Treatment planning

Introduction

Over the years, in various aspects of medical science, technology has been used to find ways to analyze medical data that can help in the accurate diagnosis of a case and generating a suitable treatment plan [1, 2].

The crucial parts of any successful endodontic procedure are a careful diagnosis, difficulty estimation, and treatment planning [3].

The American Association of Endodontists (AAE) Case Difficulty Assessment Form [4] is a standard form used to

collect data and analyze the difficulty level of a case. It includes potential risk factors that can complicate treatment and adversely affect the outcome. These factors are divided into patient considerations, diagnosis and treatment considerations, and additional considerations. This form provides a template for general dentists to objectively assess the case difficulty. The information thus obtained can be used to appraise the patient about the case difficulty and its subsequent prognosis and to communicate with a specialist if the general dentist feels the need for referral of the case owing to the difficulty.

Artificial intelligence, particularly machine learning, has significantly contributed to the decision-making process in medical practice. It has mainly been applied in medicine for diagnostics and prediction of prognosis [5]. When a machine learning system is expected to make decisions from a discrete and finite number of output classes, it is termed as a classifier [6]. The algorithm learns from the given data and derives a classifier which can be used to make decisions for new patients in the future. Machine learning algorithms are predictive

✉ Shivani Mallishery
sshivani96@gmail.com

¹ Nair Hospital Dental College, Mumbai 400008, India

² Northeastern University, Boston, USA

³ Department of Conservative Dentistry and Endodontics, Nair Hospital Dental College, Mumbai, India

and have a level of accuracy associated with their predictive quality. Even though the accuracy may not have a perfect score, they do give acceptable results provided the training is done using quality data and extracting useful features from the data.

In the field of dentistry, the use of machine learning has progressed remarkably. It has been applied to understand the factors in preventing toothache [7], to determine the effect of restorative material differences on the lifespan of a restoration [8], and to analyze the variations between dentists in the diagnoses of caries [9].

The aim of this study is to integrate machine learning decision-making model for assessing the difficulty of an endodontic case using the AAE Endodontic Case Difficulty Form.

Methodology

Using the AAE Case Difficulty Assessment Form after obtaining the patients' consent, 500 patients who visited the dental school were diagnosed as requiring endodontic treatment. The patients' dental findings and medical history were recorded after obtaining ethical clearance from the Institutional Ethics Committee (EC-84/CONS-09ND/2018). The forms were filled by the primary researcher. The filled AAE forms were assessed by two endodontists who were not a part of the study. The calibration of endodontists was done by seating them separately in a dimly lit room after being explained the purpose of the study. They were then asked to examine 5 radiovisiographs (RVGs) which were not a part of the study, and provide a referral decision. Both endodontists had 100% consensus about the 5 RVGs. After this calibration, they were asked to examine the study RVGs. In case of conflicting opinions, a third endodontist opinion was taken. Radiographic findings were recorded using radiovisiography, CR100, CleaRay (Intense Medical and Dental System Pvt. Ltd., Delhi, India). The assessment would decide whether the case difficulty warrants a referral to a specialist or not. An Android application was developed to record each filled form along with its referral decision. It provided a user interface same as that of the paper-printed version of the AAE Case Difficulty Assessment Form; this facilitated easy digitalization of the forms. The application was built in Android Studio (Google Inc., Mountain View, CA, USA), a software to build applications for the Android platform, using Java programming language. The data entered was stored on Firebase [10] Realtime Database (Google Inc., Mountain View, CA, USA). Cloud storage ensured centralized storage of data and availability of all data through a single point for the server to train the machine learning model. Firebase, a product of Google, is aimed at providing cloud storage, analysis tools, and authentication services to mobile applications. The general process

of machine learning starts with training, where the data is fed to the algorithm along with the referral decision; here the algorithm tries to learn by observing questions (filled form) and answers (referral decision). Next phase is to evaluate how well the algorithm has learnt, by testing it. During testing, the algorithm is only provided questions (filled forms) and it gives answers (referral decision) which are checked to be correct or not. Evaluation metrics are further detailed in the "Discussion" section. Figure 1 outlines the steps involved in this study.

Data pre-processing

Data for a single case is represented by a filled form and a decision regarding referral of that case. To maintain anonymity in the data, no information regarding patient identity was recorded. Pre-processing involved separating out the filled form information as the input features and the decision regarding referral as the output features. The input features are originally binary as the form contains yes/no observations. This study presents two approaches for input feature preparation; one of which uses these binary input features as they are, while the other one summarizes the data to a certain level. The data set for the first method has 83 input features, which is the total number of yes/no observations and will be referenced further as data-raw.

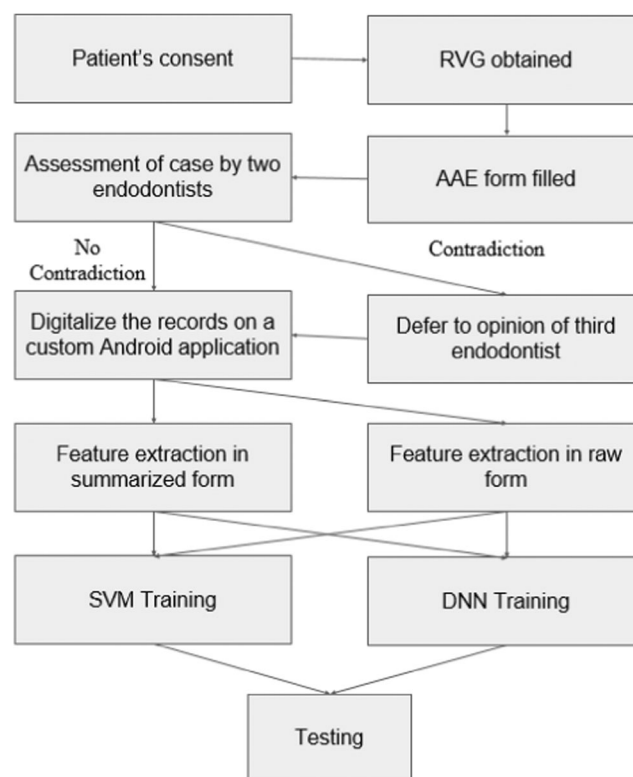


Fig. 1 Methodology of the study

The AAE form has 17 main features under consideration, and each feature has multiple yes/no observations. A summary can be obtained for getting just 17 numerical features. Each observation has an inherent difficulty level associated with it: “minimal difficulty,” “moderate difficulty,” or “high difficulty” with an assigned weightage of 1, 2, or 5 respectively. Each of the 17 features is initially assigned a value of 0, and, if any observation under some feature is present, then the corresponding weight value is added to the respective feature. This prepares the input features for all records. The following algorithm depicts this procedure for a single record. The data set generated from it will be referenced as data-summarized.

```

procedure summarizeFeatures(R) //R is the record.
  D = [] //D is the set of summarized features
  for f in R //f has observations for one feature
    v = 0 //initial value set to 0
    for o in f //o is a single observation
      w = weight(o) //finds corresponding w for o
      if o == yes then //check if o is yes
        v = v + w //add weight to feature value
      end if
    end for
    D.add(v) //add feature value v to set D
  end for
  return Dend procedure.

```

Machine learning algorithms

Taking the above two prepared data sets, two machine learning algorithms, support vector machine (SVM) [11], and deep neural network (DNN) [12] were trained with them. SVM with a linear kernel was used. A linear SVM tries to find the best linear equation (or hyperplane in case of more than two input features) separating the data points as per their respective output classification. While doing this, it tries to maximize the margin distance between the equation's line and the closest data points on both sides of the line. These closest data points on both sides are called the support vectors. The linear equation (or hyperplane) and support vectors together form an SVM model which can be used for further classification of unseen records.

DNN is a multi-layered architecture where each layer consists of multiple neurons (data processing modules) and each layer's neurons are connected to all the neurons of the next layer. They are capable of modelling highly complex functions. The proposed architecture consists of one input layer, two hidden layers, and an output layer of one neuron, which gives binary output. Input layer neurons have no special

function; they just forward, unchanged, the input feature values to the next layer. The hidden and output layer neurons have an activation function. The ReLU (rectified linear unit) activation function is used, which gives the positive part of input as output. The function is given below.

$$f(x) = \max(0, x)$$

A deep neural network attempts to minimize the error in classification; this defines its cost function. AdaGrad optimizer [13] is used to achieve this task of minimization. It features a dynamic learning rate which decreases as the algorithm approaches to a global minimum.

For training a machine learning algorithm, any dataset is randomly split up into two parts, training set and test set. This ensures that the test set, which is used for testing the performance of an algorithm, uses data which is unseen for the algorithm. If training set data was used for testing as well, the algorithm may seem to perform very well but, it may not generalize well enough for new data when putting into practical use. A more sophisticated version of this basic split of data is cross-validation, and a stratified 10-fold cross-validation [6] approach is used in training and testing of the algorithm. The randomness for splitting the dataset has an implication that every time the algorithm is trained, the split obtained will be different which will introduce a small level of variation in the results obtained, but the trend in efficiency between different algorithms will not change that easily as the variation is mostly within ~2% of any performance metric.

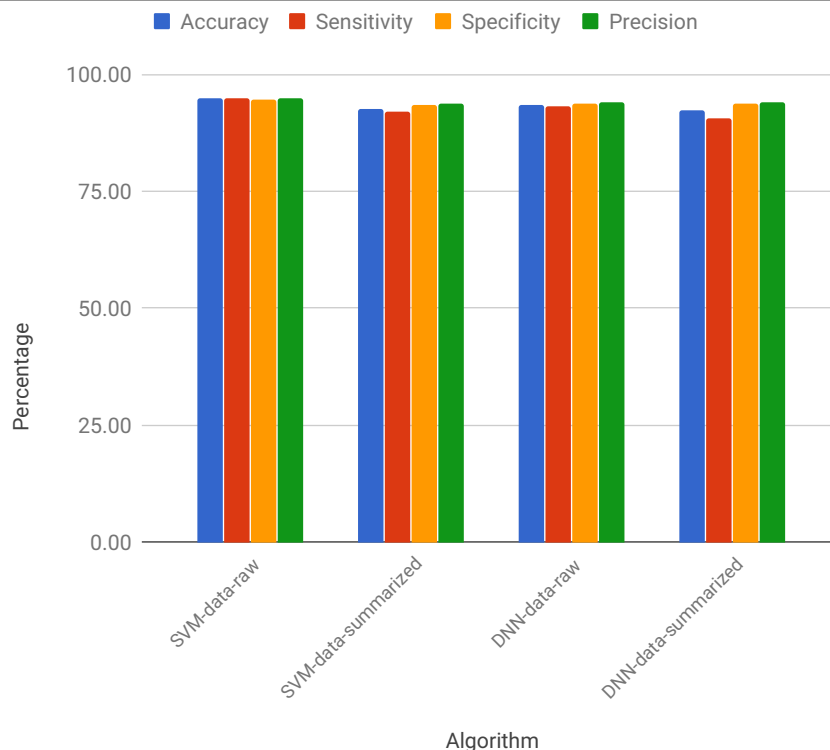
Results

There were four ML models tested in this project, which are listed out in Table 1. The metrics considered for evaluation were accuracy, sensitivity, specificity, and precision. Table 1 shows the values for each of the metrics in units of percentage. For easier comparison, Fig. 2 shows the evaluation results graphically. Table 2 shows the number of patients that were referred and a number of patients that were not referred by each candidate ML model.

Table 1 Evaluation of candidate machine learning models

Algorithm	Accuracy	Sensitivity	Specificity	Precision
SVM-data-raw	94.80	94.96	94.63	94.96
SVM-data-summarized	92.60	91.86	93.39	93.68
DNN-data-raw	93.40	93.02	93.80	94.12
DNN-data-summarized	92.20	90.70	93.80	93.98

Fig. 2 Graphical chart of evaluation of machine learning models



Discussion

Before initiating any endodontic procedure, clinicians should introspect on their capability to handle the case. If they are unable to do so, they should refer the case to a specialist who would be able to successfully complete the case [14]. Hence, the assessment of a case before initiating the treatment is mandatory. The AAE designed the Endodontic Case Difficulty Assessment Form to enable a practitioner to objectively understand the level of case difficulty. This can help the clinician to make decisions regarding the treatment that is in the patient's best interests. The form utilizes three categories of considerations to understand the complexity of the treatment [15]:

1. Patient considerations (medical history, anesthesia, mouth opening, anxiety, pain, and swelling)

2. Diagnostic and treatment considerations (radiographic difficulties, complexities in isolation of tooth, canal and tooth morphology, resorption, radiographic appearance of canals)
3. Additional considerations (history of trauma, previous endodontic treatment, periodontal condition)

Although the guidelines of AAE form provides no clear-cut demarcation for categorizing a case as minimal, moderate, or high difficulty, the AAE educator's guide [16] provides a point score to be assigned to each item within each difficulty level. Items listed in "minimal difficulty" category are assigned a point value of 1, in "moderate difficulty" category of 2, and in "high difficulty" category of 5. The following score ranges are then used to make a decision regarding referral:

- a. Less than 20 points: Dental student may treat—level of faculty supervision tailored according to the student's level of experience
- b. 20–40 points: An experienced and skilled dental student may treat with very close supervision by an endodontist, or the case be referred to an endodontist or graduate student.
- c. Above 40 points: Case should be referred to an endodontist or graduate student.

Table 2 Distribution of patients referred and not referred

	Not referred	Referred
Actual	242	258
SVM-data-raw	242	258
SVM-data-summarized	247	253
DNN-data-raw	245	255
DNN-data-summarized	251	249

Radiovisiographs were used to obtain radiographic findings. Radiovisiography was chosen as a diagnostic tool since the technique provides a quicker diagnostic procedure compared to traditional radiographic techniques. It enables archiving of the collected data. It also has an added advantage in terms of reduced radiographic exposure to the patient of approximately 95% over conventional Ektaspeed film. It allows digitalization of a conventional film and eliminates the need to process the films. Further, measurements like the angle of curvature and linear measurements can be recorded using radiovisiography [17].

As the AAE Endodontic Case Difficulty Assessment Form in this study has been assessed by two endodontists, the machine learning algorithm derives its intelligence from the decisions made by these specialists. It thus helps a general dentist to understand beforehand the complexities of the case as assessed by a specialist, so that they can avoid any endodontic failure mid-treatment and refer the patient to a specialist. Further, when doubts exist about the referral of a case to a specialist, it can serve as a guide for dental students and novice dental practitioners.

Linear model-based machine learning algorithms are the most convenient and simple algorithms to become a candidate, but their model assumption makes it heavily underfit data in most cases. Decision trees can practically fit any type of data and have no underlying model assumption, but the depth of the trees grows very fast and it can easily overfit unless some complex ensemble methods are used. Deep neural networks become a popular choice due to recent advancements in technology and specialized GPU hardware available for the complex computations involved in it. Moreover, it assumes a non-linear model and can fit any non-linear function with the choice of correct architecture. Support vector machines are powerful and have the ability to fit non-linear data by using different kernels. Hence, the last two algorithms serve to be good candidates for the given case where the underlying model of data is not guaranteed to be linear.

During the training process, the algorithm tries to find patterns in data common to a given classification (whether to refer or not). This makes the algorithm leverage onto certain areas of the form which are most prominent to determine the referral decision. These important areas are determined based on the dataset provided, purely on observation as it would be done by any individual trying to manually find patterns. This gives an advantage that if only these specific areas of the form are filled, even then the algorithm will be able to make a sound prediction. In the current dataset, the decision of referral depends more on local factors than systemic factors. Hence, even if only local factors are filled accurately, the model will be able to provide good predictions on the case.

According to the guidelines of AAE assessment form, when a single high difficulty factor, systemic or local, is present for a case, the case is automatically categorized as a high difficulty case that warrants referral to an endodontist, despite all other factors belonging to minimal difficulty. However, when the algorithm is used, the decision of referral depends more on local factors, as systemic factors belonging to the “high difficulty” category of the AAE form can be managed by a well-equipped general dentist, or by a general dentist in a hospital-based setup. For example, root canal treatment of a molar with all local factors categorized as “minimal difficulty,” in a patient with serious illness or disability, can be managed by a well-equipped general dentist. While the same case would be termed as a high difficulty case requiring referral, when the manual AAE form is used.

The results listed in the previous section analyze the alternate algorithms based on four performance metrics. Henceforth, patients referred to a specialist will be termed as positives, and patients who do not require referrals will be termed as negatives. Accuracy [6] is the percentage of correctly classified test set records. It does not give any information about the distribution of positives and negatives. Sensitivity [6] is the percentage of correctly classified positive records, and specificity [6] is the percentage of correctly classified negative records. These two metrics give a good amount of information about the positives’ accuracy and negatives’ accuracy. Referring a patient to a specialist, even when it is not necessary, might not be harmful as the patient gets better treatment. However, not referring to a specialist a patient who is supposed to be referred to is a dangerous scenario which might cause complications in further treatment. Hence, a better ML model is one that has high sensitivity. Precision [6] is the percentage of positively classified records that are actually such. This metric will support the fact that the ML model is not simply classifying every patient record as positive, irrespective of its ideal classification.

It is observed in Fig. 2 that the SVM algorithm performs better on both data sets and that the best performance is given on data-raw. Some models seem to have higher specificity compared with sensitivity, but both of these metrics are disparate. SVM-data-raw, the most highly sensitive model, is best suited for solving problems investigated in this study. In Table 2, it is seen that the actual distribution of patients referred and not referred is fairly even. So, any algorithm will not be biased towards one of the two possible predictions. It can also be observed that SVM on raw data has the same distribution as that of the actual results, but this does not mean that the same 258 patients that were supposed to be referred were predicted to be referred by that algorithm. This can be

perceived through the four performance metrics. If it were the case that the same 258 patients that were to be referred were predicted by the algorithm, then all metrics would have been 100%.

The use of machine learning along with the AAE form, therefore, provides several advantages over using the form alone:

1. Provides an expert endodontist's opinion regarding difficulty level assessment and referral of a case, which especially benefits less-experienced dentists or novice dental practitioners.
2. Although the form itself is divided into three categories of difficulty, manual difficulty level assessment becomes complicated when multiple factors from each category are present. The form aids in providing automation to this crucial step of case diagnosis.
3. Unlike the use of AAE form alone, the algorithm can also predict the referral decision when only specific areas of the form are filled.
4. The experience level of general dentists differs from one country to another. The algorithm can be trained according to the standards of a particular country.

The manual form categorizes the case as minimal, moderate, and high difficulty based on the factors present. Unlike this, the algorithm uses only two difficulty levels, which translates directly into a referral decision. However, the algorithm does not provide a clear-cut demarcation between the two categories, as it varies from case to case.

This study presents a solution which aims mainly to assist dental students and less-experienced general practitioners to assess the difficulty level of a case and to predict the need for referral by using the algorithm to prevent endodontic failure. However, the algorithm once trained cannot customize the output according to the experience of the general dentist. It relies on the data provided through the AAE form alone. The general dentist's own experience in the field of endodontics is not considered. Hence, the algorithm might fail to provide proper prediction in the case of an endodontic treatment performed by a highly experienced general dentist.

Conclusion

Endodontic failures can be caused by factors that cannot be predicted even by the most astute endodontists before initiating treatment, and are sometimes unavoidable. The AAE Endodontic Case Difficulty Assessment Form makes predictions easier. It particularly helps in case selection, to avoid any failure during or after treatment. Machine learning, when

combined along with the AAE Endodontic Case Difficulty Assessment Form, allows a dentist to refer the case to an endodontist after case selection and assessment, to avoid undue errors during the treatment. This improves the quality of endodontic treatment rendered to the patient and also prevents the need for re-treatment.

The support vector machine algorithm, a predictive machine learning (ML) algorithm, has been shown to make the best predictions pertaining to the referral of an endodontic case. The AAE Endodontic Case Difficulty Assessment form, being a standard, works as a good source of data to the algorithm. The trained algorithm can be employed in practice to assist decisions about referral of an endodontic case. Its further scope in this domain could be to investigate alternative sources of data, such as radiographic images, alone or combined with the AAE Endodontic Case Difficulty Assessment Form, to enable utilization of computer vision, with which alternative algorithms can be explored. Artificial intelligence can further be employed in the analysis of dental magnetic resonance imaging, computed tomography, and cephalometry.

Acknowledgments We thank all the participants of this study.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval The study has been approved by the appropriate Ethics Committee (Nair Hospital Dental College; Mumbai, India approval no. EC-84/CONS-09ND/2018). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Reiser SJ, Reiser SJ. (1981) Medicine and the reign of technology. Cambridge University Press
2. Foster KR, Koprowski R, Skufca JD (2014) Machine learning, medical diagnosis, and biomedical engineering research-commentary. *Biomed Eng* 13(1):94. <https://doi.org/10.1186/1475-925X-13-94>
3. Carrotte P (2004) Endodontics: part 2 diagnosis and treatment planning. *Br Dent J* 197(5):231. <https://doi.org/10.1038/sj.bdj.4811612>

4. https://www.aae.org/specialty/wp-content/uploads/sites/2/2017/06/2006casedifficultyassessmentformb_edited2010.pdf. Accessed on 8 February 2019
5. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S (2017) Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2:230–243. <https://doi.org/10.1136/svn-2017-000101>
6. Han J, Pei J, Kamber M (2011) *Data mining: concepts and techniques*. Elsevier
7. Kim EY, Lim KO, Rhee HS (2009) Predictive modeling of dental pain using neural network. *Stud Health Technol Inform* 146: 745–746
8. Käkilehto T, Salo S, Larmas M (2009) Data mining of clinical oral health documents for analysis of the longevity of different restorative materials in Finland. *Int J Med Inform* 78(12):e68–e74. <https://doi.org/10.1016/j.ijmedinf.2009.04.004>
9. Korhonen M, Gundagar M, Suni J, Salo S, Larmas M (2009) A practice-based study of the variation of diagnostics of dental caries in new and old patients of different ages. *Caries Res* 43(5):339–344. <https://doi.org/10.1159/000231570>
10. <https://firebase.google.com/>. Accessed on 25/09/2018
11. Deng N, Tian Y, Zhang C (2012) *Support vector machines: optimization based theory, algorithms, and extensions*. Chapman and Hall/CRC
12. Goodfellow I, Bengio Y, Courville A, Bengio Y (2016) *Deep learning*. MIT press, Cambridge
13. Duchi J, Hazan E, Singer Y (2011) Adaptive subgradient methods for online learning and stochastic optimization. *J Mach Learn Res* 12:2121–2159
14. Swartz DB, Skidmore AE, Griffin JA (1983) Twenty years of endodontic success and failure. *J Endod* 9(5):198–202. [https://doi.org/10.1016/S0099-2399\(83\)80092-2](https://doi.org/10.1016/S0099-2399(83)80092-2)
15. Chandra BS, Gopikrishna V (2014) *ENDODONTIC PRACTICE*. Wolters Kluwer India
16. <https://www.aae.org/specialty/wp-content/uploads/sites/2/2017/06/educatorguidetocdaf.pdf>. Accessed on 7/7/19
17. Ingle JI, Bakland LK, Baumgartner JC (eds) (2008). *Ingle's endodontics 6*. PMPH USA

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.