Review Article

A Survey of Dental Caries Segmentation and Detection Techniques

Vincent Majanga [ID] **and Serestina Viriri** [ID]

Statistics and Computer Science, University of KwaZulu-Natal, Durban 4000, South Africa

Correspondence should be addressed to Serestina Viriri; viriris@ukzn.ac.za

Dental caries detection, in the past, has been a challenging task given the amount of information got from various radiographic images. Several methods have been introduced to improve the quality of images for faster caries detection. Deep learning has become the methodology of choice when it comes to analysis of medical images. This survey gives an in-depth look into the use of deep learning for object detection, segmentation, and classification. It further looks into literature on segmentation and detection methods of dental images through deep learning. From the literature studied, we found out that methods were grouped according to the type of dental caries (proximal, enamel), type of X-ray images used (extraoral, intraoral), and segmentation method (threshold-based, cluster-based, boundary-based, and region-based). From the works reviewed, the main focus has been found to be on threshold-based segmentation methods. Most of the reviewed papers have preferred the use of intraoral X-ray images over extraoral X-ray images to perform segmentation on dental images of already isolated parts of the teeth. This paper presents an in-depth analysis of recent research in deep learning for dental caries segmentation and detection. It involves discussing the methods and algorithms used in segmenting and detecting dental caries. It also discusses various existing models used and how they compare with each other in terms of system performance and evaluation. We also discuss the limitations of these methods, as well as future perspectives on how to improve their performance.

1. Introduction

For easier segmentation and detection of dental caries, prior knowledge for several tasks is needed. There is a need to understand the various sections of the tooth and the specific position of the lesion on the tooth. An understanding of the types of dental images to be used, for instance, panoramic or bitewing radiographs, is also needed. Furthermore, the specific regions or areas of interest which are required should be clear in order to be able to choose the suitable method for segmentation and detection of caries. All this information is required in order to achieve high performance segmentation and detection of dental caries.

1.1. Tooth Anatomy. The tooth is a small white structure found in the jawbone of animals and human beings. In human beings, the number of teeth ranges from 20 primary teeth in children to 28–32 permanent teeth in adults. Further, the tooth can be broken down into three main layers:

the crown, neck, and root. The crown is the visible region above the gum line. The neck connects the crown to the root, while the root is the region inside the bone socket [1]. Additionally, each tooth consists of hard tissues that protect the soft tissues in the center, as can be seen in Figure 1.

1.2. Dental Caries. Dental caries is a tooth infection caused by bacteria. Diet that includes fructose, sucrose, and glucose accelerates the occurrence of dental caries. The acid released from the above process leads to demineralization of the tooth surface. Caries occurs when the rate of demineralization is less than the rate of decay. There are various categories of dental caries and these can be characterized [3]. The protocol characterizes dental caries based on its location and also the affected tooth [4]. The categories are as follows:

Class I: dental caries on occlusal surfaces of posterior teeth, for instance, molars and incisors.

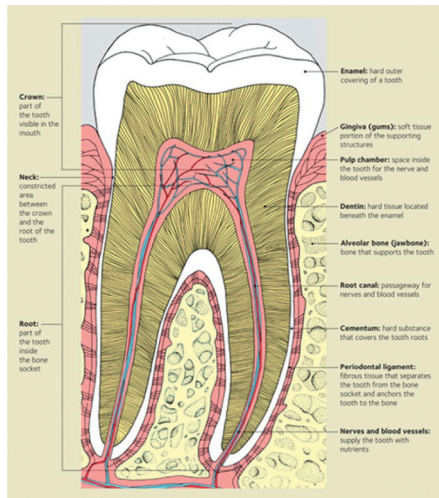Class II: occurs on proximal surfaces of posterior teeth.

FIGURE 1: Cross section of a molar tooth [2].

Class III: occurs on interproximal surfaces of anterior teeth, with no incisor edge involvement.

Class IV: occurs on interproximal surfaces of anterior teeth with incisor edge involvement.

Class V: occurs on the lingual or cervical third of the facial surface of the tooth.

Class VI: occurs on the occlusal or incisor edge, worn away due to attrition.

From positional classification, caries can also be classified based on the severity of lesions on the tooth. This is done based on the amount of dentin and enamel that has been affected by the caries.

Incipient caries are caries that have a depth of less than half of the enamel of the tooth.

Moderate caries are caries that are more than halfway through the enamel but do not touch the dentin.

Advanced caries are caries that extend to the dentin region.

Severe caries are caries that extend more than halfway through the dentin and even reach the pulp.

Identification of caries under classes I, IV, and VI can be done during clinical inspection, since the regions are visible orally. The introduction of X-rays in the medical field has greatly improved diagnosis of various ailments. In the dental field, radiography has improved the visual inspection of patient's teeth. X-rays have enabled professionals to be able to view previously unobservable regions of caries that would have gone untreated.

1.3. Dental Radiographs: X-Rays. There are varying degrees of information needed depending on the form of treatment required to diagnose a certain ailment. An X-ray or radiograph is a digital film that represents unobservable information not visible by the naked eye. Figures 2–4 show some types of radiographs. The work in [5] explains three types of X-rays that are commonly used to diagnose dental
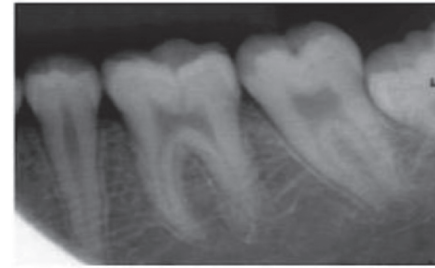

FIGURE 2: Periapical radiograph.

health. There are two types of radiographs, intraoral and extraoral [6].

Intraoral radiographs: X-ray film captures radiographic images while inside the mouth. This type is subdivided further into *bitewing radiographs*, *periapical radiographs*, and *occlusal radiographs*.

Extraoral radiographs: X-ray film captures radiographic images while outside the mouth. This type is subdivided further into *panoramic radiograph*, *computed tomography (CT)*, and *sialography*.

2. Deep Learning Systems Application Areas

The introduction of deep learning systems has contributed to several application areas using medical images.

2.1. Digital Microscopy and Pathology. Deep learning methods have been very popular in these areas, especially with the growing availability of tissue specimen. In this domain, deep learning techniques developed have had their main focus on three broad challenges. These are segmentation of large organs, detecting, segmenting, and classifying nuclei, and also detecting and classifying region of interest of lesions. Other areas where deep learning techniques have contributed include normalizing histopathology and color normalization in image analysis. The work in [7] introduces a method for normalizing stains on histopathology images using autoencoders. Color normalization is further demonstrated by [8], where the use of convolutional neural networkss (CNN) for tissue classification in stained images was introduced. *Brain.* In this area, deep neural networks (DNN) have been used for brain image analysis in several domains, most notably the classification of Alzheimer's disease. Other domains include segmentation of brain tissues and anatomical structures, for instance, hippocampus. There are also the segmentation and detection of brain tumor, lesions, and microbleeds. There are other tasks that require more anatomical information like white matter lesion segmentation, and [9] tackled such scenarios. They lowered the sampling rate of nonuniformed sampled patches to cover a larger part of the region of interest.

Chest. Deep learning also has contributed to thoracic image analysis, on both tomography and radiology. Deep neural networks have addressed the detection, characterization,
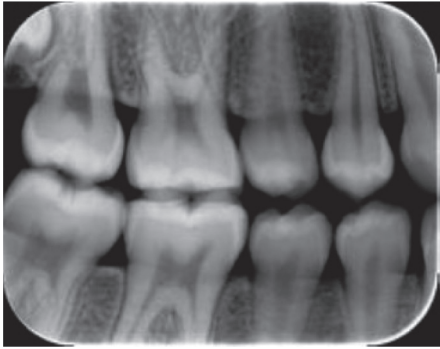
Figure 3: Bitewing radiograph.

Figure 4: Panoramic radiograph.

and classifying nodules from radiology tests conducted. Computed tomography (CT) scans have detected several diseases, including lung diseases, from a single image. Chest radiography is the most common radiology test and therefore uses a large set of images that are used to train systems. These systems can be a combination of convolution neural networks used for image analysis and recurrent neural networks used for text analysis.

Eye. Deep learning algorithms have also been introduced in the analysis of eye images and have seen CNNs being employed to address the segmentation of anatomical structures. These networks have also addressed the detection and segmentation of retinal diseases, diagnosis, and assessment of image quality. The work in [10] shows the performance of a Google Inception v3 network for diabetic retinopathy detection and compares the results with those of seven ophthalmologists.

Musculoskeletal. In this domain, deep learning algorithms are used for identification and segmentation of joints, bones, and soft tissues in images. The method used in [11] is one of the applications that trained their system with musculo-skeletal disc images and had very good performance across some radiology scoring tasks.

Breast. Research on breast imaging has resulted in [12], which shows significant advances over the state of the art, achieving performance of human readers on the region of interest (ROI). The main task also is to detect breast cancer,

and this consists of several subtasks. These include detection and classification of lesions, micro classification, and cancer risk scores of images. The availability of huge amounts of image data has made mammography easier to perform and the most common method used for breast radiography.

Abdomen. Research on the abdomen is localized on the segmentation of organs, mainly pancreas, kidneys, liver (tumors), and bladder. The main radiograph used for most organs is the CT radiograph, as well as MRI only for prostate analysis. The method used by [13] introduces a hybrid neural network used to extract features that will be used further for classification. *Cardiac.* Cardiac image analysis has embraced deep learning in segmentation, tracking, classification, and accessing image quality. The MRI is used for radiographic testing. Reference [14] introduced the use of neural networks to segment the left ventricle using the recurrent connection of U-network architecture. Reference [15] used neural networks to perform regression for identifying some cardiac sequence on its model.

3. Factors That Lead to Dental Caries

[The biology, prevention, diagnosis and treatment of dental caries Scientific advances in the United States] show factors affecting caries, namely, saliva, bacteria, diet, and hereditary.

Bacteria. Reference [16] shows how plaques form around the surface of the tooth and then come into contact with carbohydrates to form acids that dissolve the tooth structure. This concept formed the basic foundation of dental caries. There is no specific way of how bacteria affects the tooth; thus it is hard to control the dental caries disease. In the method [17], how dental caries occurs from bacteria associating with fermented carbohydrates is explained, thus being referred to as a diet-bacteria disease. Researchers have associated the caries process with fermented carbohydrates [18] and showed a relationship between sugar and caries leading to acids formed on the tooth's surface. Rather, [19] demonstrated how frequent sweet snacks consumption is related to dental caries.

Saliva. It plays a critical part in the well-being of soft and hard tissues of teeth inside the mouth [20]. When the saliva flow rate is low, this shows risk to dental caries. Further, [21] explained how saliva flow measurement is an important risk assessment and management measure for dental caries.

3.1. Prevention of Caries. Dental caries can be prevented through various ways:

> Sealants: this can be done as [22] shows how sealants are introduced over the specific carious region to stop the decay process on the tooth. This prevents food particles from collecting in the tooth pits, thus preventing caries.

> Remineralization: this is done by hardening the tooth's enamel as explained by [23] to prevent dental caries from even taking place.

Fluoride: this is found from excessive fluorosis in drinking water; in [24], a practicing dentist, associated with the enamel [3], showed its effects on the enamel of the tooth. Research from [25] gave an optimal level of fluoride in water to prevent dental caries.

Risk assessment: this entails the prior determination of someone's developing dental caries during a specific period of time as [26] explained to enable easier management. Assessment is very useful in determining whether extra diagnostic measures are required. The work in [27] explained how assessment assesses effectiveness of previous caries control measures and acts as a guide to treatment planning in the future. Further research has seen that caries can be prevented by saliva when fluoride application is combined with regular removal of forming dental plaques, realized by brushing teeth. The research in [28] explained how saliva and small amounts of fluoride contributed to the hardening of the enamel and thus low risk of dental caries.

3.2. Diagnosis of Caries.

3.2. Diagnosis of Caries. Diagnosis of dental caries is done by the following:

Clinical method: this is described by [29] and is explicitly defined as the visual detection of dental caries from an oral examination. They used separators to visualize areas of interest and also used dental floss applied on these areas to detect roughness of the surfaces.

Radiographic method: the radiographs also known as X-rays are digital images and [30] introduced them in dentistry. The work in [31] further intensified the use of dental radiographs in their research to detect caries. The work in [32] generated images of higher diagnostic quality compared to those found from conventional films to detect caries in bitewing X-rays.

3.3. Treatment of Caries. Focus has shifted from surgical ways to the management of caries via restoration of tooth structure development or implants. The research work in [33] emphasized the prevention of the disease, remineralization steps, and minimizing the access to caries affected regions to avoid further decay.

4. Image Representation for Dental Segmentation and Detection

The representation of dental images is done through the segmentation of various regions of interest from the larger image to locate objects. Segmentation of images is, therefore, partitioning of an image into several segments to be used to identify objects and their edges. Image segmentation can be categorized according to similarity and discontinuity properties. Discontinuity based methods are referred to as boundary-based methods, while similarity-based methods are referred to as region-based methods [34].

Therefore, the segmentation process is based on dividing an image into groups of similar characteristics and features. Mathematically, segmentation of an image R is a finite set of regions $R_1 \ldots R_s$. $R = \cup i = 1 R_i R_i \cap R_{ji} \neq j$.

4.1. Categories of Dental Images Segmentation. Research done by [35, 36] further categorizes segmentation methods according to various characteristics such as region, entropy, shape, threshold, and pixels correlation among others. These characteristics were from thermal, X-ray images to aid analysis of specific points or regions of interest. Research studies show that dental image segmentation is classified as region-based, cluster-based, threshold-based, boundary-based, and watershed-based methods.

Region-based: it divides an image into several regions based on discontinuities of pixel intensity. Reference [37] explained how segmentation of dental panoramic X-ray images assists dentists to detect osteoporosis disease. The work in [38] also used the region-based approach in segmentation of bitewing dental X-ray images.

Threshold-based: this is done by choosing a threshold value from pixel intensities of an image. Then, pixels that exceed the threshold value are placed into a region, while those that are below the threshold value are placed in an adjacent region. Threshold-based segmentation is common in most of the reviewed articles [2, 6, 39–48].

Cluster-based: it is the automatic grouping of image data based on certain degrees of similarity between the data. The degree of similarity depends on the problem being solved. The algorithm used to perform clustering of data uses the automatically detected groups as initial parameters. Research by [49] performed jaw lesion segmentation using the fuzzy C-means method. The work in [50] used a semifuzzy supervised clustering algorithm to segment dental radiographs.

Boundary-based: it is used to find edge or point discontinuities on images. It detects color or pixel intensity discontinuities in the gray levels of the image. Active contours are used by [51–53], as one of the approaches to segment images based on their boundaries. The approach performs segmentation by outlining an object from an image and is also referred to as the snake method. Level set method (LSM) is another approach for detecting boundaries in an image. It handles segmentation by performing geometric operations to detect contours with topology changes. Examples of works on boundary-based segmentation are [54, 55], which used LSM to segment radiographic images.

Watershed-based: it is performed on a grayscale image and used mathematical morphology to segment adjacent regions in an image; watershed-based segmentation was used by [56] on bitewing dental radiographs. It was also used by [57] as a combination of the K-means

clustering and the watershed method for color-based segmentation.

4.2. Diagrammatic Representation of Dental Image Segmentation.

This section aims at giving pictorial representations of various segmentation methods, as can be viewed below in Figure 5. The work in [38] proposed a novel way of finding region of interests in both the gap valley and tooth isolation using edge intensity curves. It used the region growing approach [58] to detect the region of interest. It further used canny edge detection algorithm [59] to detect the edges of the isolated teeth.

A dental classification method by [60] proposed a classification of periapical images via the fuzzy value, which takes care of dental orientation problems such as missing teeth. They also used the dental universal numbering system to categorize teeth into incisors, canine, molar, and premolar as shown in Figure 6.

Segmentation of the images is done through multiscale aggregation [61], which deals with pixel distortions in image data. Integral projection is used to detect horizontal individual teeth, and this is further classified according to one of the four incisor, molar, premolar, and canine categories. Figure 7 shows images of the above-mentioned method.

Additionally, [62] proposed the use of faster regions of a convolution neural network to detect and number periapical dental images. A filtering algorithm is used to delete overlapping boxes detected by the faster convolution network. A neural network is introduced to detect missing teeth. A rule-based teeth numbering system is used to match labels of detected teeth boxes to modify results that violate set rules. The work in [63] explained how they achieved instance segmentation through the use of a mask region-based convolution neural network. This system is an extension of [64], which includes a section of convolution networks to achieve the task of instance segmentation (Figure 8).

Features are extracted from ResNet101, which then compose the feature pyramid network (FPN). The FPN defines anchors and extracts regions of interest (ROI). The region proposal network is formed from a combination of anchors and the feature pyramid network (FPN) (Figure 8). Finally, regions of interest are aligned to the same size and, further, each fixed size is classified as a tooth or a background (class scores). The fixed size features are localized by regression of bounding box coordinates. Finally, pixels are segmented by the full convolution network (FCN) [23] in each detected mask as seen in Figure 9.

Table 1 clearly shows related works grouped by various categories and methods. A diagnostic system proposed by [66] consists of Laplacian filtering, window-based adaptive threshold, morphological operations, statistical feature extraction, and back-propagation neural network. The back-propagation neural network is used to classify a tooth surface as normal or having dental caries. A study by [67] analyzes feature extraction performance of dental caries image using Gray Level Cooccurrence Matrix (GLCM) algorithm for contrasted two types of caries based on the theory of GV Black, namely, dental caries Class 3 and Class 4.

A CNN model using a U-shaped deep CNN (U-Net) proposed by [68] was for dental caries detection on bitewing radiographs, and it was further investigated whether the model can improve clinicians' performance.

5. Feature Extraction Techniques for Dental Images

Feature extraction on an image is done by the use of various methods depending on its texture, pixels, and color intensity. The method by [69] proposed evaluating the performance of various texture feature maps to recognize demineralization of caries. It used intraoral image analysis that includes run-length matrices (RLM), first-order features (FOF), cooccurrence matrices, gray tone matrices, local binary patterns, and K-means clustering to transform images of confirmed caries cases. The performance from the different feature maps was compared to that of radiographic images by several radiologists. Feature maps are a product of extraction of features from an image by various methods and techniques that include the following.

First-order features (FOF): this is a formalised description of the detection of probability of a particular intensity within data in an image, and this shape is used to determine image parameters such as contrast, sharpness, and other objects. For instance, for an image I with spatial resolution $W * H$ and range intensity G, its histogram is defined as

$$H(i) = \frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} \begin{Bmatrix} 1, & I(x, y) = i \\ 0, & \text{otherwise} \end{Bmatrix} i \in [0..G-1]. \quad (1)$$

Several equations can be derived from the above equation, namely, mean, variance, and entropy, and can be represented as

$$FOF_{\text{mean}} = \frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} I(x, y), \quad (2)$$

$$FOF_{\text{variance}} = \frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} \left(I(x, y) - FOF_{\text{mean}}\right)^2, \quad (3)$$

$$FOF_{\text{entropy}} = \frac{1}{WH} - \sum_{i=1}^{G} H(i) \log(H(i) + \epsilon). \quad (4)$$

Run-length matrix is defined by changes in the illumination of pixel values on an image [70], to explain the coarse texture of pixels can be expressed by a larger section of similar color encoded as a matrix. Other works that have embraced this method include [71–73].

Gray level cooccurrence matrix (GLCM): this, according to [74], shows texture relations between adjacent pixels in the image texture. The matrix is calculated from entries that represent probabilities of coexisting gray tone pixels that are next to each other, and the distance between the pixels is a
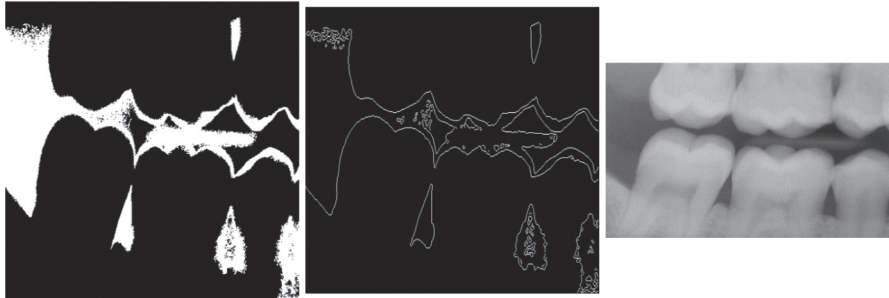
FIGURE 5: Left: region growing approach; central: canny edge detected image; right: tooth isolation and gap valley using binary intensity integral curves [38].
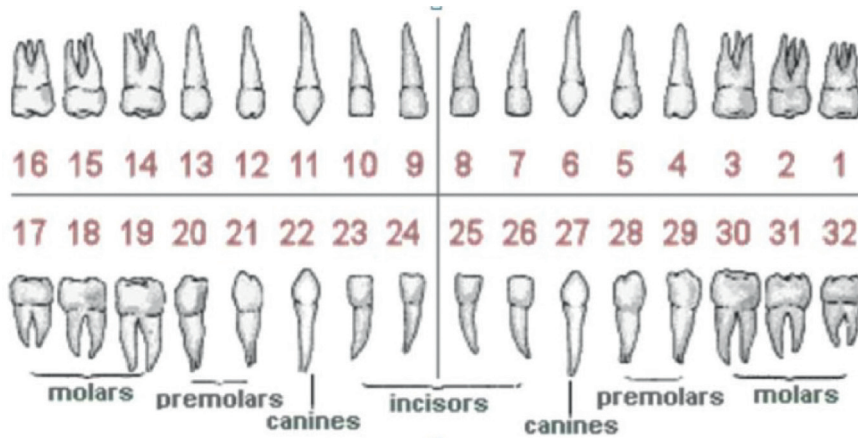

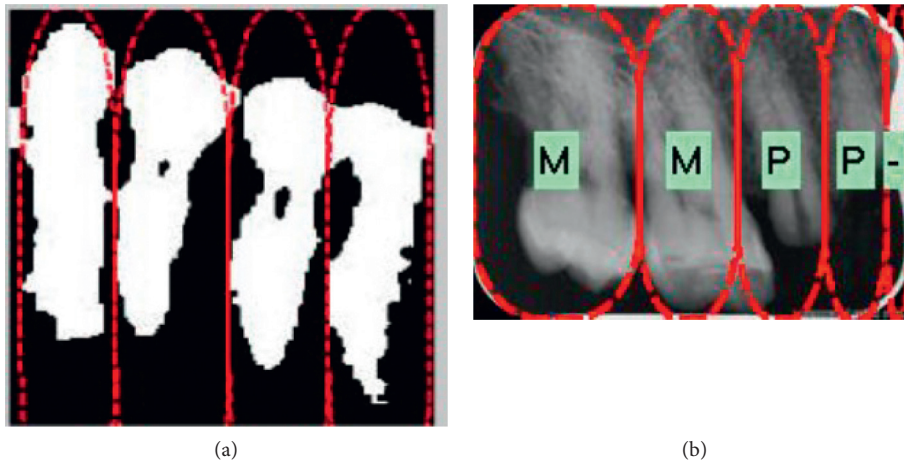FIGURE 6: Dental universal numbering system [60].

FIGURE 7: (a) Horizontal integral projection of lower tooth; (b) classification result [61].

feature parameter. Other features that can be extracted by this matrix include contrast, entropy, variance, some average, and homogeneity. Related works include [75, 76].

Gray tone difference matrix (GTDM): this matrix, according to [77], is described as gray tone textural properties, which is the difference in pixel intensity level I and illumination I in a $K * K$ neighborhood described as

$$\text{GLTDM}(i) = \frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} |i - \overline{I}| \text{ where } \overline{I} = \frac{1}{M-1}, \quad (5)$$

$$\sum_{m=-K}^{K} \sum_{n=-K}^{K} I(x+m, y+n), (m,n) \neq (0,0). \quad (6)$$
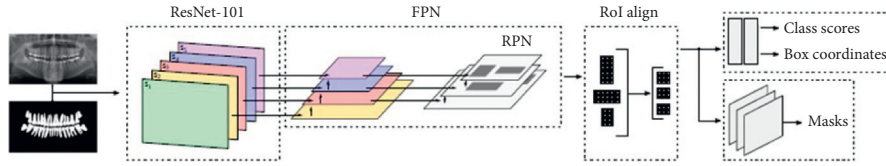
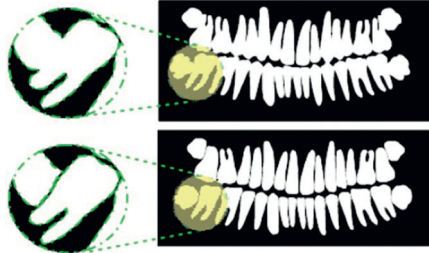FIGURE 8: Training process of the segmentation system [63].

FIGURE 9: Process of separating the teeth.

TABLE 1: Related works grouped by various categories and methods.

Category	Method	Related works
Region-based	Region growing approach	[37, 38]
Threshold-based	Histogram-based threshold	[2, 6, 39–47, 65]
Cluster-based	Fuzzy C-means	[49, 50]
Boundary-based	Active contours, level set method	[51–55]
Watershed-based	Watershed	[56, 57]

Another method proposed by [78] combines GTDM, LBP, and K-means clustering for feature extraction.

Laws' texture energy measures: according to [79], a method that utilised set masks to calculate local energy of an image was proposed. These masks further detect textural characteristics such as edges, spots, levels, and ripples. Other works that use this method include [80, 81].

Local binary patterns (LBP): it detects small structures such as edges, lines, and spots [82], on the skin, thus representing them as binary patterns. For pixel inputs (x_c, y_c), a neighborhood of radius R and several evenly sampled points on radius P are specified. Local binary pattern function is given by

$$LBP_{P,R}(x_{c,y_c}) = \sum_{p=0}^{P-1} s\left[I(x_c, y_c) - I(x_p, y_p)\right].2^p. \qquad (7)$$

Here, s[.] operator returns 1 for positive values and 0 for negative values.

Figure 10 shows a $3 * 3$ local binary operation concatenating all 8 bits to output a binary number, which is converted to a decimal number. This decimal number is the LBP code and is assigned to the center pixel. Other works that use LBP include [83] and [84].

Clustering: this used K-means method to cluster pixels. The intensity of pixels is taken into account; also prior information about the number of clusters should be defined. Euclidean distance is the metric used when adding individual pixels to the nearest cluster. A cluster value is set to the average value of all pixel intensities, and this procedure is iterated until a specified threshold is met [85]. Related works include [86–88]. Figure 11 shows the feature maps of the extraction techniques discussed.

According to [89], a textural feature system for diagnosis of dental caries in radiographs was used. This system introduced other feature extraction techniques such as the following.

Gabor filters: they were originally introduced by Gabor, D. (1946) and extracted edge-like components with very high frequency in a local region of an image. They are described as the best texture descriptor due to their use in segmentation, object recognition, tracking of motion, and image registration. Spatial domain Gabor is defined as

$$g(x, y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right)\exp\left[\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi jWx\right]. \qquad (8)$$

Here σ_x and σ_y are standard deviations for (x, y) axis distribution, and the sinusoidal frequency is denoted by W. Other related studies on Gabor filters include [90–92].

Local ternary patterns: this, from [93], shows how LBP is extended to a three valued code from two values. The ternary code is got from comparing neighboring pixel values with a set threshold value τ. Values that are within the threshold value are set to 0, those above it are set to +1, and those below the threshold value are set to −1.

The threshold function value is defined as

$$f(x_i, x_c, \tau) = \begin{cases} 1, & \text{if } x_i \geq x_c + \tau, \\ 0, & \text{if } |x_c - x_i| < \tau, \\ -1, & \text{if } x_i \leq x_c - \tau, \end{cases} \qquad (9)$$

where τ is the set threshold, x_c is the value of the central pixel, and x_i are the neighbouring pixels of x_c. Diagrammatically, the process can be displayed by a Figure 12 LTP code for a 3×3 matrix image region. Other related works include [94–96].

Morphological gradient: the method introduced by [97] is used to increase the intensity of boundary edges of an image. This method makes it easier to observe edge boundaries and other objects clearly [98] and identify dental caries on teeth. Other works that involve the mMG method include [99–101]. Multiple morphological gradient consists of several encryption elements that aid in processing of images and include gradient, multiple values, and threshold.

Morphology gradient: it increases the intensity of the edges of objects in the image, and this is done by dilation subtracted by erosion morphology: $(A \oplus B) - (A \ominus B)$.

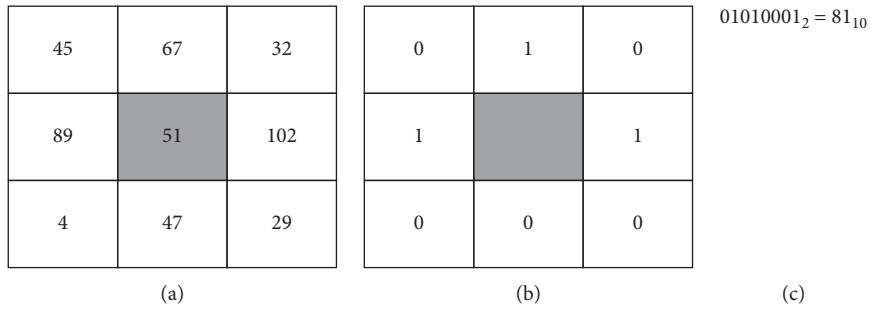| 45 | 67 | 32 | | 0 | 1 | 0 | | $01010001_2 = 81_{10}$ |
| 89 | 51 | 102 | | 1 | | 1 | | |
| 4 | 47 | 29 | | 0 | 0 | 0 | | |

| (a) | (b) | (c) |

FIGURE 10: LBP operation given $P = 1$ and $R = 1$ [82].
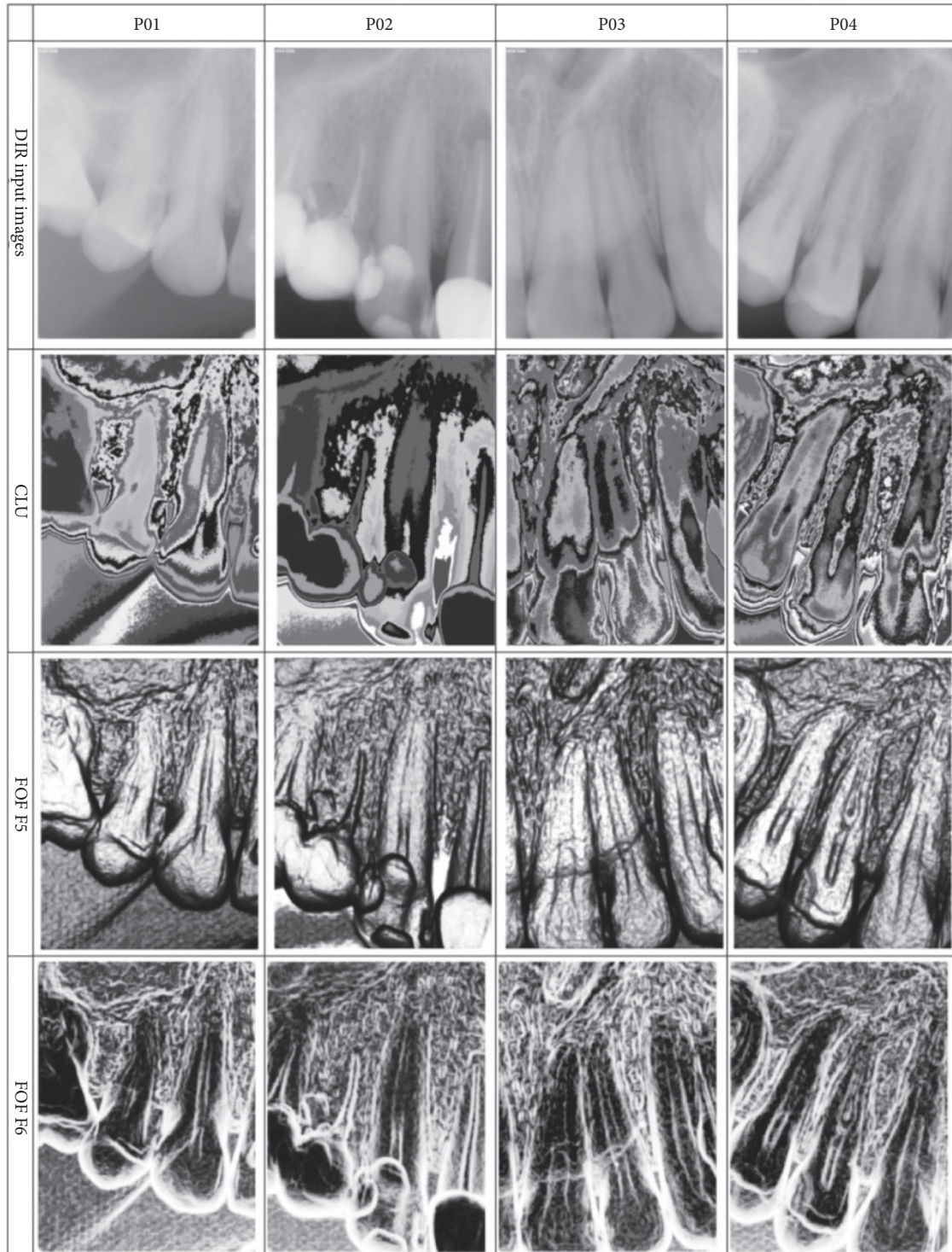
FIGURE 11: Intraoral radiographs texture feature maps [69].

85	32	71
53	50	47
64	38	15

1	−1	1
0	x	0
1	−1	−1

0	1	0
0	x	0
0	1	1

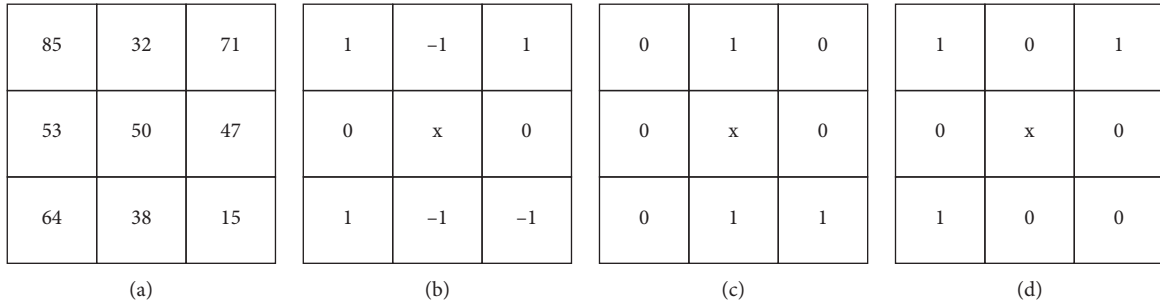| 1	0	1
0	x	0
1	0	0

(a) (b) (c) (d)

FIGURE 12: (a–d) LTP code with $\tau \geq +15, \tau \leq -15$ and both positive and negative LBP codes.

Multiplier value: it is a constant value to enable the increase in pixel value on the image. This is defined as

$$mv = \frac{\max(\max(A))}{w}. \quad (10)$$

Here w is the bit depth.

Threshold: this is done by separating pixel values into two classes, 0 and 1, which are dependent on a constant threshold value q. According to [99], thresholding is defined as

$$T(A) = \begin{cases} 1, & A < q, \\ 0, & A \geq q. \end{cases} \quad (11)$$

Here $0 < q \leq A_{\max}$.

Table 2 shows various works grouped by feature extraction method.

A novel deep convolution layer network (CNN) with a Long Short-Term Memory (LSTM) model was proposed by [103] for the detection and diagnosis of dental caries on periapical dental images.

6. Deep Learning Methods and Algorithms

Research by [104] explains how machine learning algorithms are divided into unsupervised and supervised learning algorithms.

6.1. Learning Algorithms. Supervised learning: the model is represented with a dataset $D = \{x, y\}_{n=1}^{N}$ of input features x and y. They can take different forms depending on the learning task, for instance, with classification y will be a scalar representing a class label. With regression, it will take a vector of continuous variables. In dealing with a segmentation model, y can be a multidimensional label image; basically supervised training finds model parameters θ that best predict data given a loss function $L(y, \hat{y})$. This $\hat{y}$ describes the output of the model obtained by feeding x to $f(x; \ominus)$ that represents the model.

Unsupervised learning: there are algorithms that are trained to find patterns, such as data without labels, for instance, clustering and principal component analysis methods. These algorithms can be performed on models with different loss functions.

TABLE 2: Works grouped by feature extraction method.

Feature extraction method	Related works
FOF	[69]
Run-length matrix	[70–73]
LBP	[82–84]
Clustering	[85–88]
Laws of texture energy measures	(Laws, K., 1980) [80, 81]
GLCM	[74–76]
GTDM	[77, 78]
Gabor filters	[90–92, 102]
LTP	[93–96]
Multiple morphology gradient (mMG)	[98–101]

6.2. Neural Networks. These are networks that contribute to deep learning systems. Neural networks consist of neurons with some activation a and parameters $\theta = \{W, \beta\}$, where W is weight and β is bias. Activation preselects the linear combination of input x to a neutron and its parameters, followed by element on element (.) for nonlinearity, referred to as the transfer function: $a = \sigma(w^T + b)$. Transfer functions for traditional neural networks are sigmoid and the hyperbolic tangent function. Multilayered neural networks known as perceptrons consist of several layers of these transformations:

$$f(x; \theta) = \sigma\left(W^L \theta\left(W^{L-1} \ldots \sigma\left(W^0 x + b^0\right) + b^{L-1}\right) + b^L\right). \quad (12)$$

Here W^n is a matrix comprising w_k rows with k as activation on the output. n is the number of current layers, and L is the final layer. Hidden layers are layers in between the input layer and the output layer. When a network contains multiple hidden layers, it is referred to as a deep neural network.

Activations done to the final layer which is the output layer of the network are mapped to a distribution over the number of classes $P(y|x; \theta)$ through a softmax function:

$$P(y|x; \theta) = \text{softmax}(x; \theta) = \frac{e^{(W_i^L)^T x + (b_i)^L}}{\sum_{k=1}^{K} e^{(W_k^L)^T x + (b_k)^L}}. \quad (13)$$

Here W_i^L is the weight of the output node associated with class i.

Stochastic gradient descent is the method used to fit parameters to dataset D. A small subset of the dataset is used

to optimize maximum likelihood and this in return minimizes negative log-likelihood as

$$\text{argmin}_\theta - \sum_{n=1}^{N} \log[P(y_n|x_n;\theta)]. \tag{14}$$

This further leads to binary cross entropy loss for two class problems and categorical cross entropy for multiclass tasks. Deep neural networks gained popularity in 2006 and, in the method in [105], it was explained how DNNs were pretrained layer by layer and their stacked network fine-tuned to produce good evaluation performance. Currently, the most used deep learning popular networks are the convolution and the recurrent networks.

6.2.1. Convolution Neural Networks. These networks have weights that are shared in a manner that the network performs convolutions on images. This means that there is no redundancy in the way the model learns separate detectors for the same object that occurs at different position on an image. It reduces the number of parameters to be used for learning. At each layer of the network, the image is convolved with a set of K kernels $W = (W_1 \ldots W_K)$, with biases $\beta = (b_1 \ldots b_K)$ and each generating X_k feature map.

The features are subjected to a nonlinear transform $\sigma(.)$ and the process iterated for every convolution layer l given by

$$X_K^1 = \sigma\left(W_K^{l-1} * X^{l-1} + b_k^{l-1}\right). \tag{15}$$

The convolution neural network also has pooling layers that aggregate pixels and their neighbors using an invariant function, which is the maximum operation. Aggregation is done to subsequent convolution layers and, at the end of the stream, a regular network layer is added where weights are not shared. The network is trained by feeding activations to the output layer through a softmax function. Under the CNNs, there exist common architectures that are widely used in the analysis of medical images and include the following.

General classification architectures: they include networks such AlexNet [106], which, unlike its precursor LeNet [107], consists of five convolution kernel layers employed in the input and output. AlexNet incorporated rectified linear units (ReLU) as their activation function, and this has become the most common choice in CNNs. Further, there has been an interest in using smaller kernels instead of single layers of kernels with a large receptive field; this in return has less number of parameters. Research by [108] discussed the use of deeper neural networks that have small fixed size kernels in each layer of the 19-layer model, referred to as VGG19 network model. More complex building blocks have been introduced to deep networks, to improve efficiency in the training process, and also reduce the amount of parameters used. There is also [109], where GoogLeNet is a 22-layered network that uses inception blocks with a set of convolutions of different sizes. The ResNet architecture was introduced and consists of ResNet blocks which, instead of learning a function, learn mappings in each layer that is close

to the identity function [64]. There has been an increase in the use of these deep architectures due to their low memory use, and this even contributed to a recent version of GoogLeNet referred to as Inception v3 [110].

Multistream architectures: they are networks that accommodate multiple inputs in form of channels towards the input layer and then are later merged at any point in the network. Image processing can be done by the use of multiscale image analysis and classification for brain lesions [111]. Multistream architecture [112] is used for segmentation of natural images. Challenges of deep learning systems in the medical imaging domain are in adapting existing architectures, for instance, with different input formats (2D or 3D) data. Volumetric data can be divided into slices and fed as different streams to a network to avoid a result of large amounts of parameters. These techniques can still be used to perform knee cartilage segmentation [113]. The multistream architectures can be used for classification in the context of medical imaging [114].

Segmentation architectures: they are specific to the task of segmentation of medical images. CNNs are used to classify individual pixels with those in their neighborhood in an image. To avoid redundancy in the classification of the pixels, fully connected layers are rewritten as convolutions and this helps the CNN to take in input images larger than what it was trained on and produce a likelihood map. The resultant fully CNN (fCNN) can then be applied to an entire volume of images. This further leads to low resolution of output compared to the input images due to pooling of layers. There is a technique that applies the FCNN to shifted versions of the input, followed by stitching the result together to obtain a full resolution of the final output [115]. The FCNN was improved by proposing the U-net architecture that has convolutions in its downsampling and later an upsampling task to increase the image size [116]. Another method added skip connections to U-net architecture to connect downsampling and upsampling of the convolution layers [115]. A similar approach was used by [117], for 3D data, and [118] incorporated residual blocks and Dice loss layer to the U-net architecture instead of the commonly used cross entropy.

6.2.2. Recurrent Neural Networks. These networks were developed for discrete sequence analysis and have varied lengths for both inputs and outputs, thus making them suitable for tasks such as machine translation. In a classification task, the model learns distribution over classes $P(y|x_1 \ldots x_T;\theta)$ given sequence $x_1 \ldots x_T$ as input.

The RNN has a hidden state h at time t which is the output of a nonlinear mapping from x_t and previous state h_{t-1}: $h_t = \sigma(Ex_t + Rh_{t-1} + b)$. Here W, R are weight matrices that are shared over time. For a classification task, several fully connected layers are used and followed by a softmax to map the sequence over the classes as

$$P(y|x_1 \ldots x_T;\theta) = \text{softmax}(h_T: W_{\text{out}}, b_{\text{out}}). \tag{16}$$

RNNs have problems of memory shortage similar to those of other deep neural networks. Several techniques have

been developed, such as the Long Short-Term Memory (LSTM) cell [119], to deal with problem. A simplification of the LSTM is the gated recurrent unit (GRU) [120]. RNNs are increasingly being adopted with promising results, for instance, [121], in the human brain challenge.

6.3. Unsupervised Models. Some unsupervised models reviewed include autoencoders, restricted Boltzmann machines, and deep belief networks.

6.3.1. Autoencoders. These are networks trained to reconstruct input x on output y via a hidden layer h. They have a weight matrix $W_{x,h}$ and bias $b_{x,h}$ from input to hidden state and $W_{h,x}$ with corresponding bias $b_{h,x}$ from hidden state to reconstruction. The hidden activation is computed as

$$h = \sigma\left(W_{x,h}x + b_{x,h}\right). \tag{17}$$

Here, h is smaller than x to prevent the model from learning the identity function. According to [122], another solution was introduced to prevent the model from learning the identity function. The model uses a denoising autoencoder that trains the model to reconstruct input from noise. Deep autoencoders are realized by placing autoencoder layers on top of each other, and in most cases the layers were trained individually. Examples of autoencoders include the following.

Variational autoencoders: with these, [123] introduced the use of conditional variational autoencoders for pathology detection in medical images.

6.3.2. Adverserial Networks. These are used for image generation tasks and include works from [124, 125].

6.3.3. Deep Belief Networks. These are a type of Markov random field (MRF) which constitute an input layer $x = (x_1 \ldots x_N)$ and a hidden layer $h = (h_1 \ldots h_M)$ having a latent feature representation. Its connections are bidirectional and thus a generative model that can be sampled for new data points. An energy function can be defined for a state (x, h) of input and hidden layers as

$$E(x, h) = h^T W_x - c^T x - b^T h. \tag{18}$$

Here c and b are the bias terms. Further, the probability of the system is given by

$$p(x, h) = \frac{1}{Z}\exp\{-E(x, h)\}, \tag{19}$$

This makes computing the partition function Z intractable, while conditioned inference in computing h conditioned on x is tractable and is given by

$$P\left(h_j|x\right) = \frac{1}{1 + \exp - b_j - W_j x}. \tag{20}$$

The use of DBNs can be to fuse medical images [48]. DBNs also are used to extract high level features from medical images and effectively classify them [126].

6.4. Software and Hardware. These are the processors and software programs used to handle the running of various deep learning techniques. They include GPU computing libraries such as CUDA and OpenCL, these being very fast processor units compared to the previously used CPUs. These GPUs work hand in hand with the available open-source software programs that provide a platform to implement various operations of neural networks such as convolutions. Most popular packages include Caffe [127], TensorFlow [128], Theano [129], and Torch [130] that provide interfaces for implementing various operations in deep learning. They are also third-party packages written on top of frameworks like Keras [131] and GitHub [132]. Figure 13 shows a representation of deep learning architectures.

7. Deep Learning Uses in Medical Images

These systems can be used in various tasks like segmentation, classification, detection, and registration.

7.1. Classification. In this approach, one has a single input or multiple inputs with a single variable as output. For instance, in a disease classification setup, one has the disease or not and diagnosis of the disease is based on a sample of the dataset. Transfer learning is therefore realized and is defined as the use of pretrained networks to cover very large datasets for deep networks training. Transfer learning can be used to fine-tune a pretrained network on medical images and also for feature extraction on image data. These processes are beneficial in saving time used to train deep networks and enable extracted features to be analyzed faster. Object classification is usually done on a small part of the image and can be divided into two or more classes for analysis. For instance, [133] used three CNNs, each taking a nodule patch as input and each feature output is concatenated to form a final feature vector. Multistream CNNs are used to classify skin lesions, with each stream working on different resolutions of the image [134].

7.2. Detection. Detection of objects and their respective regions of interest in images is an important part of diagnosis by medical practitioners. The task consists of localization and identification of small regions in a full image. These systems are designed to automatically detect regions and decrease the reading time of human experts. An example of detection is [98], which explained the use of edge detection method to detect dental caries on dental images.

7.3. Segmentation. Segmentation of various organs allows deep analysis of several parameters related to volume and shape, for example, in skin or breast cancer analysis. It is always the first step before the detection process and is defined as the identification of the pixels which make up the interior or exterior contour of the object of interest. There have been a wide variety of methods developed to segment images using deep learning medical images. Segmentation is
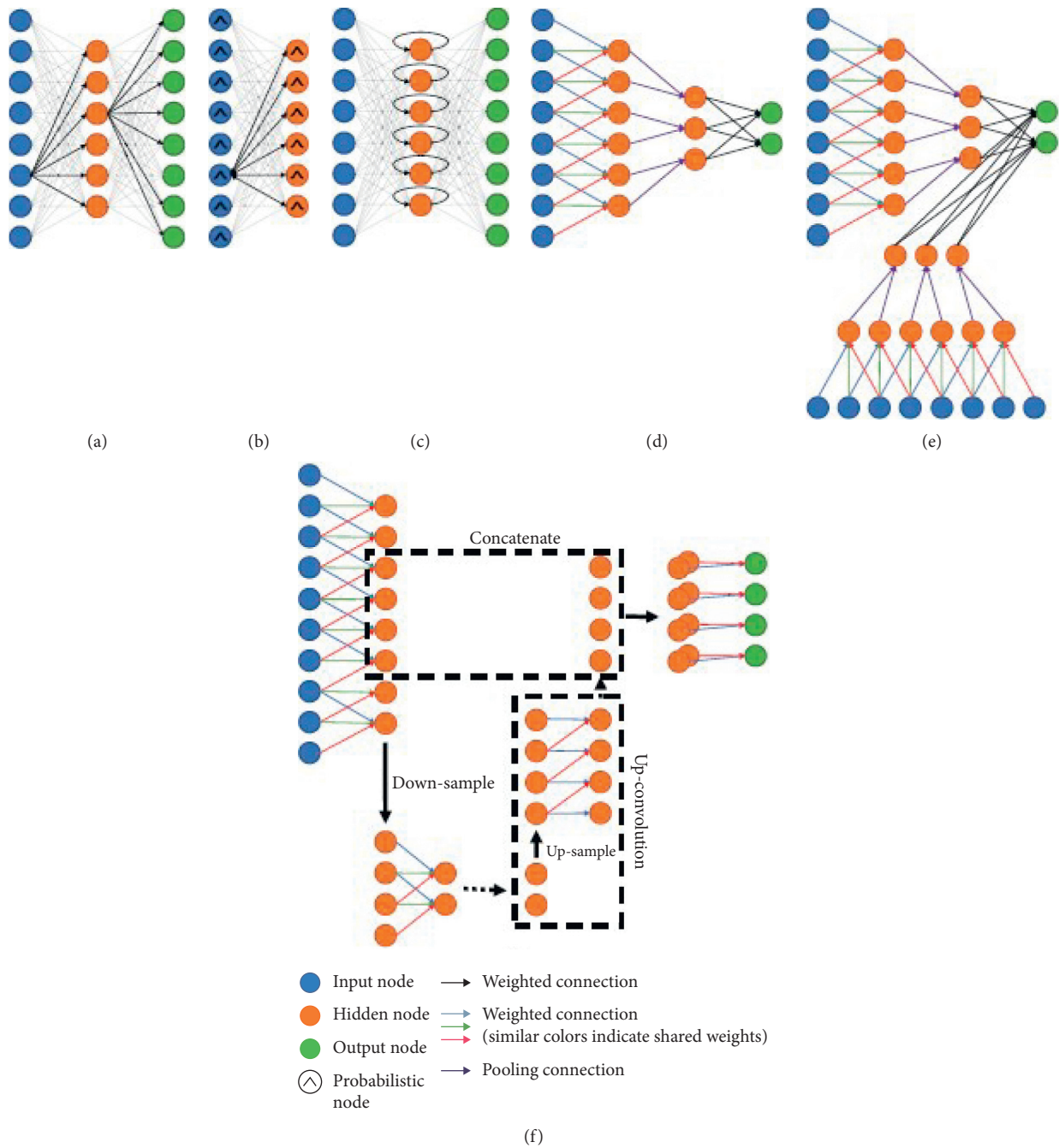
(a)   (b)   (c)   (d)   (e)

(f)

FIGURE 13: Representation of deep learning architectures. (a) Autoencoder. (b) Deep belief networks. (c) Recurrent neural network. (d) Convolutional neural network. (e) Multistream convolutional neural network. (f) U-net architecture [104].

needed to perform accurate segmentation on 3D CNNs using multistream networks with different scales [135].

7.4. Registration. Registration is a common image analysis task in which coordinate transforms are calculated from one image to another. This is done iteratively assuming parametric transformations and a predetermined metric is optimized. Deep networks can benefit from registration by estimating a similarity measure for two images to drive an iterative optimization, for instance, [136]. They can also be

used to predict transformation parameters using regression, for example, [137].

8. Databases Used by Dental Images

There is a need for efforts to build or come up with a public dataset to aid in developing of algorithms to be used in the dental imaging area. In order for this to come to pass, researchers need to release data used in their papers, and this will lead to a repository that can reliably catalogue and archive publicly dental imaging data. Several examples of

datasets exist and include the digital database for screening mammography (DDSM) [138], which is used for mammogram image analysis and aids in screening of breast cancer. There exist several databases that are used in dental imaging and these include the following:

PASCAL VOC 2007: this is from the results from the pascal visual object classes challenge of 2007 [139].

Caltech101: this is a dataset of images to facilitate computer vision and its techniques [140]. This dataset is also applicable in image recognition and classification and contains a total of 9,146 images split into 101 distinct object categories.

NORB: it is used for experiments in 3D object recognition shapes [141]. The dataset contains 50 toys belonging to five generic categories that include aeroplanes, cars, four legged animals, and trucks.

CIFAR-10/100: this consists of 60000 32×32 color images in 10 classes, with 50000 training images and 10000 test images [142].

MNIST: it is an acronym for modified national institute of standards and technology, and it is a dataset of handwritten digits commonly used for training and testing data in the machine learning field. It contains 60000 training images and 10000 testing images [143].

LabelMe: it is a dataset and web based tool used for image annotation [144] and was created by MIT computer science intelligence laboratory (CSAIL). It is applicable in computer vision and is very dynamic, free, and open to public contribution. It contains 187240 images, 62197 annotated images, and 658992 labelled objects.

ImageNet: it is a large dataset used for visual object recognition and has more than 20000 categories with each category containing several hundred images [145].

Summary: both MNIST and CIFAR-10/100 datasets are available for the public, while the other datasets can be accessed by directly contacting the researchers. Dental datasets are difficult to find and thus researchers prefer using available public datasets. ImageNet is one of the datasets that is used for evaluations by several dental imaging models like ResNets and VGGNets.

9. Evaluation Protocols for Dental Images

Deep learning techniques are used on problems having very large datasets with thousands of instances, and therefore they need a way to estimate the performance of a given data configuration and use this for comparison with performances of other configurations. One of the ways is splitting data [146], since very large datasets require long training times.

9.1. Splitting Data. The data is split into training and testing data splits; for instance, Keras library for deep learning provides two ways of handling the splitting of data. It can split your data into a validation set and evaluate the performance of your model on that validation set. This is done by setting the validation split argument on the fit function to a certain percentage of your training dataset such as 30% for validation.

9.2. Manual K-Fold Cross Validation. This is used as the standard evaluation method for machine leaning techniques. This method splits the dataset into k-subsets and trains the model on all the subsets one after the other except one subset that is left out as the validation set. Evaluation is done on the left-out subset, which is the validation set, and the performance is averaged across all models created. The cross-validation method is applicable on small deep learning models and is used with 5 or 10 folds. We also have the method in [147] as the training-testing split method.

9.3. Training-Testing Split. Data is split into two parts, training set and testing set. A model is then fitted to the training set, and then the fit model is used to make predictions on the testing set. This is further used to evaluate the skill of the predictions and thus referred to as the training-testing split. Training-testing split is used as an estimate of how well the model performs on a dataset, especially when presented with new data. This method is preferred, especially with very large data and slow model to train. This is because the skill score for the model is noisy due to randomness of data. The randomness of data makes the model flexible but makes it less stable; for instance, you get different results from training the same model. This can be controlled by introducing a random seed and repeating experiments multiple times. The use of random seed is basically just using the same randomness every time the model is being fit and evaluated. Get an average of the estimated model skill after running the experiments multiple times. We also have [148] that also gives a description of the confusion matrix.

9.4. Confusion Matrix. It is an $N * N$ matrix with N number of classes being predicted and is used mostly with class output models. The matrix consists of several metrics that include the following:

Accuracy: is the total number of correct predictions.

Precision or positive predictive value: is a proportion of positive correct predictions.

Negative predictive value: is a proportion of negative correct predictions.

Recall or sensitivity: is a proportion of actual positive cases correctly identified.

Specificity: is a proportion of actual negative cases correctly identified.

9.5. F1-Score. This is defined as the harmonic mean of the precision and recall values for a classification problem task.

The F1-score is given by

$$F_1 = \left\{ \frac{\text{recall}^{-1} + \text{precision}^{-1}}{2} \right\}^{-1} = 2.\frac{\text{precision.recall}}{\text{precision} + \text{recall}}. \tag{21}$$

Harmonic mean is preferred over the arithmetic mean because it takes care of extreme values, for instance, those in a binary classification model. The F1-score can also be adjusted to increase effectiveness by adding β as an adjustable parameter to get

$$F_\beta = \left(1 + \beta^2\right).\frac{\text{precision.recall}}{\left(\beta^2.\text{precision}\right) + \text{recall}}. \tag{22}$$

10. Conclusion

From this survey, various techniques, methods, and approaches have been discussed concerning the segmentation and detection of dental images. Works that stem from the industry and academia have been mentioned and discussed, which include existing algorithms, segmentation and detection methods, databases, and various protocols for evaluating performance.

There is a huge potential for use of dental radiography and especially work focused on detection of dental images. Most of the existing systems dwell much on dental segmentation and not on feature extraction (detection). There is a need to improve existing dental detection systems, and one way to do so is by the introduction of automatic blob detection technique. Blob detection has been used in other fields of medical imaging but has not seen substantive use in the field of dental imaging. The use of such image analysis techniques to determine the presence of caries aims to create a system that takes a human diagnostic approach, whereby dental caries are diagnosed based on visual interpretation of teeth.

11. Future Work

There are issues and encouraging future perspectives of study which have popped out from our discussions here and they are highlighted as follows:

Data availability and reliability: Deep learning networks require large amount of data to be able to achieve meaningful and effective performance results. Due to the nature of dental images, there is a need for hybrid datasets to aid good performance of the networks. There is a need for publicly available datasets for dental images to make deep learning in the field possible.

Data standardization: Many methods discussed here are handling the preprocessing step through manual methods, such as cropping the region of interest on an image. These methods contribute to the loss of some key details from the images. Some networks end up dividing a whole image into subregions, and this slows down the learning process that occurs one subregion after the other. There are methods like downsampling which might lead to deletion of important details and

this seems to have been due to limitations in computational power. Deep learning approaches have seen increased learning on whole images rather than the need for manual manipulation of images at the preprocessing stage in order to get more general and accurate results.

Weight regularization methods: Deep learning networks can also be improved by introducing weight regularization to improve their performance. The regularization of weights involves optimization of model hyperparameters such as the learning rate and the dropout rate. Optimization of models at several key processes, such as the use of dropout at the segmentation stage, could improve the final result, and future work will look at how to implement such optimizations. Basically, weight regularization methods are introduced into networks for parameter optimization. Furthermore, introduction of early stopping regularization technique will also help in reducing overfitting, which is always a major problem with existing models.

Hybrid approaches: Deep neural networks can also be achieved by combining several models or methods to form hybrid networks that will improve overall evaluation performance. The combination can be in any stage of the model, for instance, combining two preprocessing techniques to come with a single one to enhance image quality. This combination can also be handled by joining various attributes of different models to form one hybrid model that will enhance training, extraction, detection, and classification of objects. Combination of different convolution networks to form one hybrid network will be a good area to explore. This will save the long training and testing times that come with large networks having many convolution networks.

Oral healthcare follow-ups from medical practitioners are very important for risk assessment and management of dental caries [149–153].

Data Availability

Deep learning networks require large amount of data to be able to achieve meaningful and effective performance results. Due to the nature of dental images, there is a need for hybrid datasets to aid good performance of the networks. There is a need for publicly available datasets for dental images to make deep learning in the field possible

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] A. Waugh and A. Grant, *Ross and Wilson Anatomy and Physiology in Health and Illness E-Book*, Elsevier Health Sciences, Amsterdam, Netherlands, 2014.

[2] L. Wang, F. Shi, G. Li et al., "Segmentation of neonatal brain MR images using patch-driven level sets," *NeuroImage*, vol. 84, pp. 141–158, 2014.

[3] G. V. Black, "Operative dentistry," vol. 2, Medico-Dental Publishing Company, Chicago, IL, USA, 1955.

[4] D. Osterloh and S. Viriri, "Unsupervised Caries Detection in Non-standardized Bitewing Dental X-Rays," *Advances in Visual Computing*, in *Proceedings of the International Symposium on Visual Computing*, pp. 649–658, Berlin, Germany, 2016.

[5] M. G. Newman, H. Takei, P. R. Klokkevold, and F. A. Carranza, *Carranza's Clinical Periodontology*, Elsevier health sciences, Amsterdam, Netherlands, 2011.

[6] S. V. Tikhe, A. M. Naik, S. D. Bhide, T. Saravanan, and K. P. Kaliyamurthie, "February. Algorithm to identify enamel caries and interproximal caries using dental digital radiographs," in *Procdings of the 2016 IEEE 6th International Conference on Advanced Computing (IACC)*, pp. 225–228, IEEE, Bhimavaram, India, 2016.

[7] A. Janowczyk, A. Basavanhally, and A. Madabhushi, "Stain normalization using sparse autoencoders (StaNoSA): application to digital pathology," *Computerized Medical Imaging and Graphics*, vol. 57, pp. 50–61, 2017.

[8] A. Vahadane, T. Peng, A. Sethi et al., "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 8, pp. 1962–1971, 2016.

[9] M. Ghafoorian, N. Karssemeijer, T. Heskes et al., "April. Non-uniform patch sampling with deep convolutional neural networks for white matter hyperintensity segmentation," in *Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 1414–1417, IEEE, Prague, Czech Republic, June 2016.

[10] V. Gulshan, L. Peng, M. Coram et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.

[11] J. Mohamed, "Mechanisms of diabetes-induced liver damage: the role of oxidative stress and inflammation," *Sultan Qaboos University Medical Journal*, vol. 16, no. 2, pp. e132–141, 2016.

[12] T. Kooi, A. Gubern-Merida, J.-J. Mordang et al., "A Comparison Between a Deep Convolutional Neural Network and Radiologists for Classifying Regions of Interest in Mammography," *Breast Imaging*, vol. 23, pp. 51–56, 2016.

[13] Q. Dou, L. Yu, H. Chen et al., "3D deeply supervised network for automated segmentation of volumetric medical images," *Medical Image Analysis*, vol. 41, pp. 40–54, 2017.

[14] R. P. Poudel, P. Lamata, and G. Montana, "Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation," in *Reconstruction, Segmentation, and Analysis of Medical Images*, pp. 83–94, Springer, Berlin, Germany, 2016.

[15] B. Kong, Y. Zhan, M. Shin, T. Denny, and S. Zhang, "Recognizing End-Diastole and End-Systole Frames via Deep Temporal Regression Network," in *Medical Image Computing and Computer-Assisted Intervention,* Springer, Berlin, Germany, 2016.

[16] W. D. Miller, "The presence of bacterial plaques on the surface of teeth and their significance," *Dent Cosmos*, vol. 44, pp. 425–446, 1902.

[17] W. H. Bowen, D. Birkhed, L. Granath, and W. D. McHugh, "Dental caries: dietary and microbiology factors. systemized prevention of oral disease: theory and practice," 1986.

[18] R. M. Stephan, "Intra-oral hydrogen-ion concentrations associated with dental caries activity," *Journal of Dental Research*, vol. 23, no. 4, pp. 257–266, 1944.

[19] R. L. Weiss and A. H. Trithart, "Between-meal eating habits and dental caries experience in preschool children," *American Journal of Public Health and the Nations Health*, vol. 50, no. 8, pp. 1097–1104, 1960.

[20] ID Mandel and S. Wotman, "The salivary secretions in health and disease," *Oral Sciences Reviews*, vol. 8, pp. 25–47, 1976.

[21] M. Navazesh, "How can oral health care providers determine if patients have dry mouth?" *The Journal of the American Dental Association*, vol. 134, no. 5, pp. 613–618, 2003.

[22] S. O. Griffin, E. Oong, W. Kohn et al., "The effectiveness of sealants in managing caries lesions," *Journal of Dental Research*, vol. 87, no. 2, pp. 169–174, 2008.

[23] C. Hazirbas, L. Ma, C. Domokos, and D. Cremers, "November. fusenet: incorporating depth into semantic segmentation via fusion-based cnn architecture," in *Asian Conference on Computer Vision*, pp. 213–228, Springer, Berlin, Germany, 2016.

[24] F. S. McKay, "The relation of mottled enamel to caries read before the Chicago dental society, April 17, 1928," *The Journal of the American Dental Association (1922)*, vol. 15, no. 8, pp. 1429–1437, 1928.

[25] H. T. Dean, P. Jay, and J. W. Knutson, "Studies on Mass Control of Dental Caries through Fluoridation of the Public Water Supply," *Public Health Reports (1896–1970)*, vol. 65, no. 43, pp. 1403–1408, 1950.

[26] E. Reich, A. Lussi, and E. Newbrun, "Caries-risk assessment," *International Dental Journal*, vol. 49, no. 1, pp. 15–26, 1999.

[27] M. Fontana and D. T. Zero, "Assessing patients' caries risk," *The Journal of the American Dental Association*, vol. 137, no. 9, pp. 1231–1239, 2006.

[28] T. Koulourides, "Implications of remineralization in the treatment of dental caries," 1986.

[29] G. V. Black, "Mottled teeth: an endemic developmental imperfection of the enamel of the teeth heretofore unknown in the literature of dentistry," *Dental Cosmos*, vol. 58, pp. 129–156, 1916.

[30] D. T. Zero, M. Fontana, E. A. Martínez-Mier et al., "The Biology, Prevention, Diagnosis and Treatment of Dental Caries," *The Journal of the American Dental Association*, vol. 140, pp. 25S–34S, 2009.

[31] H. R. Raper, "Practical Clinical Preventive Dentistry Based Upon Periodic Roentgen-Ray ExaminationsA reply to Dr. Kells' paper, The Limitations of the Roentgen-Ray in Dental Practice, given, in substance, in a discussion presented before the American Dental Association, Dallas, Texas, November 10–14, 1924," *The Journal of the American Dental Association (1922)*, vol. 12, no. 9, pp. 1084–1100, 1925.

[32] A. Wenzel, "Bitewing and digital bitewing radiography for detection of caries lesions," *Journal of Dental Research*, vol. 83, pp. 72–75, 2004.

[33] J. Mirza and G. Robertson, "Vital guide to Dental implants," *Vital*, vol. 6, no. 1, pp. 19–22, 2008.

[34] R. B. Subramanyam, K. P. Prasad, and B. Anuradha, "Different image segmentation techniques for dental image extraction," *International Journal of Engineering Research in Africa*, vol. 4, no. 7, pp. 173–177, 2014.

[35] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 146–166, 2004.

[36] G. Silva, L. Oliveira, and M. Pithon, "Automatic segmenting teeth in X-ray images: trends, a novel data set, benchmarking and future perspectives," *Expert Systems with Applications*, vol. 107, pp. 15–31, 2018.

[37] A. Lurie, G. M. Tosoni, J. Tsimikas, and F. Walker, "Recursive hierarchic segmentation analysis of bone mineral density changes on digital panoramic images," *Oral surgery, oral medicine, oral pathology and oral radiology*, vol. 113, no. 4, pp. 549–558, 2012.

[38] C. K. Modi and N. P. Desai, "May. A simple and novel algorithm for automatic selection of ROI for dental radiograph segmentation," in *Proceedings of the 2011 24th Canadian Conference on Electrical and Computer Engineering (CCECE)*, pp. 000504–000507, Ontario, Canada, 2011.

[39] A. Abaza, A. Ross, and H. Ammar, "November. Retrieving dental radiographs for post-mortem identification," in *Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP)*, pp. 2537–2540, IEEE, Cairo Egypt, November 2009.

[40] A. Ajaz and D. Kathirvelu, "April. Dental biometrics: computer aided human identification system using the dental panoramic radiographs," in *Proceedings of the 2013 International Conference on Communication and Signal Processing*, pp. 717–721, IEEE, Melmaruvathur, India, 2013.

[41] A. D. C. S. Azevedo, N. Z. Alves, E. Michel-Crosato, M. Rocha, R. Cameriere, and M. G. H. Biazevic, "Dental age estimation in a Brazilian adult population using Cameriere's method," *Brazilian Oral Research*, vol. 29, no. 1, pp. 1–9, 2015.

[42] Y. Wang, C. L. Lin, and J. D. Miller, "Improved 3D image segmentation for X-ray tomographic analysis of packed particle beds," *Minerals Engineering*, vol. 83, pp. 185–191, 2015.

[43] Y. Xi, Y. Jin, B. De Man, and G. Wang, "High-kVp assisted metal artifact reduction for X-ray computed tomography," *IEEE Access*, vol. 4, pp. 4769–4776, 2016.

[44] Y. Y. Amer and M. J. Aqel, "An efficient segmentation algorithm for panoramic dental images," *Procedia Computer Science*, vol. 65, pp. 718–725, 2015.

[45] M. Lin, J.-H. Chen, X. Wang, S. Chan, S. Chen, and M.-Y. Su, "Template-based automatic breast segmentation on MRI by excluding the chest region," *Medical Physics*, vol. 40, no. 12, p. 122301, 2013.

[46] R. Indraswari, A. Z. Arifin, D. A. Navastara, and N. Jawas, "September. Teeth segmentation on dental panoramic radiographs using decimation-free directional filter bank thresholding and multistage adaptive thresholding," in *Proceedings of the 2015 International Conference on Information Communication Technology and Systems (ICTS)*, pp. 49–54, IEEE, Surabaya, Indonesia, 2015.

[47] M. R. M. Razali, N. S. Ahmad, R. Hassan, Z. M. Zaki, and W. Ismail, "December. Sobel and canny edges segmentations for the dental age assessment," in *Proceedings of the 2014 International Conference on Computer Assisted System in Health*, pp. 62–66, IEEE, 2014.

[48] M. Kaur and D. Singh, "Fusion of medical images using deep belief networks," *Cluster Computing*, vol. 34, pp. 1–15, 2019.

[49] M. K. Alsmadi, "A hybrid Fuzzy C-Means and Neutrosophic for jaw lesions segmentation," *Ain Shams Engineering Journal*, vol. 9, no. 4, pp. 697–706, 2018.

[50] T. M. Tuan, "Dental segmentation from X-ray images using semi-supervised fuzzy clustering with spatial constraints," *Engineering Applications of Artificial Intelligence*, vol. 59, pp. 186–195, 2017.

[51] A. Işın, C. Direkoğlu, and M. Şah, "Review of MRI-based brain tumor image segmentation using deep learning methods," *Procedia Computer Science*, vol. 102, pp. 317–324, 2016.

[52] U. A. A. Niroshika and R. G. N. Meegama, *Active Contour Model for the Detection of Sharp Corners in Image Boundaries*, 2013.

[53] A. Hasan, F. Meziane, R. Aspin, and H. Jalab, "Segmentation of brain tumors in MRI images using three-dimensional active contour without edge," *Symmetry*, vol. 8, no. 11, p. 132, 2016.

[54] A. Rad, M. S. Mohd Rahim, A. Rehman, A. Altameem, and T. Saba, "Evaluation of current dental radiographs segmentation approaches in computer-aided applications," *IETE Technical Review*, vol. 30, no. 3, pp. 210–222, 2013.

[55] R. Kumar, F. A. Talukdar, N. Dey et al., "Histogram thresholding in image segmentation: a joint level set method and lattice Boltzmann method based approach," in *Information Technology and Intelligent Transportation Systems*, Springer, Berlin, Germany, 2017.

[56] Y. Li, H. Shi, L. Jiao, and R. Liu, "Quantum evolutionary clustering algorithm based on watershed applied to SAR image segmentation," *Neurocomputing*, vol. 87, pp. 90–98, 2012.

[57] I. Garg and B. Kaur, "March. Color based segmentation using K-mean clustering and watershed segmentation," in *Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 3165–3169, IEEE, Noida, India, 2016.

[58] H. T. Yau, Y. K. Lin, L. S. Tsou, and C. Y. Lee, "An adaptive region growing method to segment inferior alveolar nerve canal from 3D medical images for dental implant surgery," *Computer-Aided Design and Applications*, vol. 5, no. 5, pp. 743–752, 2008.

[59] M. Sharifi, M. Fathy, and M. T. Mahmoudi, "April. A classified and comparative study of edge detection algorithms," in *Proceedings of the International Conference on Information Technology: Coding and Computing*, pp. 117–120, IEEE, Las Vegas, NV, USA, 2002.

[60] M. L. Tangel, C. Fatichah, F. Yan et al., "June. Dental classification for periapical radiograph based on multiple fuzzy attribute," in *Proceedings of the 2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS)*, pp. 304–309, IEEE, Edmonton, Canada, June 2013.

[61] M. L. Tangel, C. Fatichah, C. Fatichah, M. R. Widyanto, F. Dong, and K. Hirota, "Multiscale image aggregation for dental radiograph segmentation," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 16, no. 3, pp. 388–396, 2012.

[62] H. Chen, K. Zhang, P. Lyu et al., "A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films," *Scientific Reports*, vol. 9, no. 1, pp. 3840–11, 2019.

[63] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, and L. Oliveira, "October. Deep instance segmentation of teeth in panoramic X-ray images," in *Proceedings of the 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 400–407, IEEE, 2018.

[64] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, pp. 91–99, 2015.

[65] G. Kaur and J. Rani, "Mri Brain Tumor Segmentation Methods-A Review," 2016.

[66] V. Geetha, K. S. Aprameya, and D. M. Hinduja, "Dental caries diagnosis in digital radiographs using back-propagation neural network," *Health Information Science and Systems*, vol. 8, no. 1, pp. 8–14, 2020.

[67] Y. Jusman, R. I. Tamarena, S. Puspita, E. Saleh, and S. N. A. M. Kanafiah, "August. Analysis of features extraction performance to differentiate of dental caries types using gray level Co-occurrence matrix algorithm," in *Proceedings of the 2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, pp. 148–152, Penang, Malasyia, 2020.

[68] S. Lee, S. I. Oh, J. Jo, S. Kang, Y. Shin, and J. W. Park, "Deep learning for early dental caries detection in bitewing radiographs," 2021.

[69] R. Obuchowicz, K. Nurzynska, B. Obuchowicz, A. Urbanik, and A. Piórkowski, "Caries detection enhancement using texture feature maps of intraoral radiographs," *Oral Radiology*, vol. 36, no. 3, pp. 275–287, 2020.

[70] M. M. Galloway and G. Mm, *Texture Analysis Using Gray Level Run Lengths*, 1975.

[71] S. Eqbal, R. K. Paliwal, and R. L. Yadava, "FCM and SVM based medical image feature extraction using wavelet transform," *Tathapi with ISSN 2320-0693 is an UGC CARE Journal*, vol. 19, no. 23, pp. 24–32, 2020.

[72] J. R. F. Raj, K. Vijayalakshmi, and S. Kavi Priya, "Medical image denoising using multi-resolution transforms," *Measurement*, vol. 145, pp. 769–778, 2019.

[73] H. U. Deshmukh and S. S. Thorat, "Tumor detection from brain mri image using neural network approach," 2018.

[74] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, vol. 6, pp. 610–621, 1973.

[75] S. K. Ps and V. S. Dharun, "Extraction of texture features using GLCM and shape features using connected regions," *International Journal of Engineering and Technology*, 2016.

[76] R. Korchiyne, S. M. Farssi, A. Sbihi, R. Touahni, and M. T. Alaoui, "A combined method of fractal and GLCM features for MRI and CT scan images classification," 2014.

[77] M. Amadasun and R. King, "Textural features corresponding to textural properties," *IEEE Transactions on systems, man, and Cybernetics*, vol. 19, no. 5, pp. 1264–1274, 1989.

[78] S. Fekri-Ershad, *Innovative Texture Database Collecting Approach and Feature Extraction Method Based on Combination of Gray Tone Difference Matrixes, Local Binary Patterns, and K-Means Clustering*, arXiv preprint arXiv: 1803.04125, 2018.

[79] K. Laws, "Textured image segmentation," Ph. D. Dissertation, University of Southern California, Los Angeles, CA, USA, 1980.

[80] M. Orooji, M. Alilou, S. Rakshit et al., "Combination of computer extracted shape and texture features enables discrimination of granulomas from adenocarcinoma on chest computed tomography," *Journal of medical imaging (Bellingham, Wash.)*, vol. 5, no. 2, Article ID 024501, 2018.

[81] P. S. Ramapraba, M. P. Chitra, and M. Prem Kumar, "Effective lesion detection of colposcopic images using active contour method," 2017.

[82] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[83] M. Verma and B. Raman, "Center symmetric local binary co-occurrence pattern for texture, face and bio-medical image retrieval," *Journal of Visual Communication and Image Representation*, vol. 32, pp. 224–236, 2015.

[84] L. Liu, S. Lao, P. W. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Median robust extended local binary pattern for texture classification," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1368–1381, 2016.

[85] E. Min, X. Guo, Q. Liu, G. Zhang, J. Cui, and J. Long, "A survey of clustering with deep learning: from the perspective of network architecture," *IEEE Access*, vol. 6, pp. 39501–39514, 2018.

[86] F. Li and M. Liu, "Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks," *Computerized Medical Imaging and Graphics*, vol. 70, pp. 101–110, 2018.

[87] S. Edwards, S. Brown, and M. Lee, "Automated 3-D tissue segmentation via clustering," *Journal of Biomedical Engineering and Medical Imaging*, vol. 5, no. 2, p. 08, 2018.

[88] K. Zhao, L. Zhou, P. Qian et al., "A transfer fuzzy clustering and neural network based tissue segmentation method during PET/MR attenuation correction," *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 7, pp. 1491–1497, 2019.

[89] V. Geetha and K. S. Aprameya, "Textural analysis based classification of digital X-ray images for dental caries diagnosis," *Int J Eng Manuf (IJEM).*, vol. 9, no. 3, pp. 44-45, 2019.

[90] S. Avinash, K. Manjunath, and S. S. Kumar, "August. An improved image processing analysis for the detection of lung cancer using Gabor filters and watershed segmentation technique," in *International Conference on Inventive Computation Technologies (ICICT)*, vol. 3, pp. 1–6, IEEE, 2016.

[91] H. Aguirre-Ramos, J. G. Avina-Cervantes, I. Cruz-Aceves, J. Ruiz-Pinales, and S. Ledesma, "Blood vessel segmentation in retinal fundus images using Gabor filters, fractional derivatives, and Expectation Maximization," *Applied Mathematics and Computation*, vol. 339, pp. 568–587, 2018.

[92] M. Hegde and P. A. Sivakumar, "Development and implementation of vlsi reconfigurable architecture for gabor filter in medical imaging application," *International Journal of Engineering and Management Research (IJEMR)*, vol. 8, no. 3, pp. 71–76, 2018.

[93] X. Xiaoyang Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1635–1650, 2010.

[94] S. Murala and Q. M. Jonathan Wu, "Local ternary co-occurrence patterns: a new feature descriptor for MRI and CT image retrieval," *Neurocomputing*, vol. 119, pp. 399–412, 2013.

[95] G. Deep, L. Kaur, and S. Gupta, "Local mesh ternary patterns: a new descriptor for MRI and CT biomedical image indexing and retrieval," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, no. 2, pp. 155–169, 2018.

[96] R. Rabidas, A. Midya, A. Sadhu, and J. Chakraborty, "March. Benign-malignant mass classification in mammogram using edge weighted local texture features," *Medical Imaging 2016: Computer-Aided Diagnosis*, vol. 9785, Article ID 97851X, 2016.

[97] J. Naam, J. Harlan, S. Madenda, and E. P. Wibowo, "Identification of the proximal caries of dental x-ray image with multiple morphology gradient method," *International Journal of Advanced Science, Engineering and Information Technology*, vol. 6, no. 3, pp. 343–346, 2016.

[98] J. Naam, J. Harlan, S. Madenda, and E. P. Wibowo, "The algorithm of image edge detection on panoramic dental X-ray using multiple morphological gradient (mMG) method," *International Journal of Advanced Science,*

Engineering and Information Technology, vol. 6, no. 6, pp. 1012–1018, 2016.

[99] Y. Liang, L. Sun, W. Ser et al., "Hybrid threshold optimization between global image and local regions in image segmentation for melasma severity assessment," *Multidimensional Systems and Signal Processing*, vol. 28, no. 3, pp. 977–994, 2017.

[100] A. Rakhlin, A. Shvets, V. Iglovikov, and A. A. Kalinin, "Deep Convolutional Neural Networks for Breast Cancer Histology Image Analysis," in *Lecture Notes in Computer Science*, pp. 737–744, Springer, Cham, 2018.

[101] S. M. Naqi, M. Sharif, M. Yasmin, and S. L. Fernandes, "Lung nodule detection using polygon approximation and hybrid features from CT images," *Current Medical Imaging*, vol. 14, no. 1, pp. 108–117, 2018.

[102] D. Gabor, "Theory of communication. Part 1: the analysis of information," *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, vol. 93, no. 26, pp. 429–441, 1946.

[103] P. Singh and P. Sehgal, "G.V Black dental caries classification and preparation technique using optimal CNN-LSTM classifier," *Multimedia Tools and Applications*, vol. 80, no. 4, pp. 5255–5272, 2021.

[104] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[105] H. Larochelle, D. Erhan, A. Courville, J. Bergstra, and Y. Bengio, "June. An empirical evaluation of deep architectures on problems with many factors of variation," in *Proceedings of the 24th International Conference on Machine Learning*, pp. 473–480, Oregon, USA, 2007.

[106] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 21, pp. 1097–1105, 2012.

[107] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[108] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, https://arxiv.org/abs/1409.1556.

[109] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," 2014, https://arxiv.org/abs/1412.6572.

[110] P. Khosravi, E. Kazemi, M. Imielinski, O. Elemento, and I. Hajirasouliha, "Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images," *EBioMedicine*, vol. 27, pp. 317–328, 2018.

[111] K. Kamnitsas, C. Ledig, V. F. J. Newcombe et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.

[112] C. Couprie, C. Farabet, L. Najman, and Y. LeCun, "Indoor semantic segmentation using depth information," 2013, https://arxiv.org/abs/1301.3572.

[113] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep Feature Learning for Knee Cartilage Segmentation Using a Triplanar Convolutional Neural Network," *Advanced Information Systems Engineering*, in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 246–253, Berlin, Heidelberg, 2013.

[114] A. A. A. Setio, F. Ciompi, G. Litjens et al., "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1160–1169, 2016.

[115] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, Boston, MA, USA, 2015.

[116] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Lecture Notes in Computer Science*, vol. 24, pp. 234–241, 2015.

[117] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016*, vol. 42, pp. 424–432, 2016.

[118] F. Milletari, N. Navab, and S. A. Ahmadi, "October. V-net: fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision,* vol. 3DV, pp. 565–571, IEEE, 2016.

[119] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[120] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014.

[121] M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber, "Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation," *Advances in Neural Information Processing Systems*, vol. 43, pp. 2998–3006, 2015.

[122] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P. A. Manzagol, and L. Bottou, "Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, no. 12, 2010.

[123] H. Uzunova, S. Schultz, H. Handels, and J. Ehrhardt, "Unsupervised pathology detection in medical images using conditional variational autoencoders," *International journal of computer assisted radiology and surgery*, vol. 14, no. 3, pp. 451–461, 2019.

[124] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "May. Self-attention generative adversarial networks," in *Proceedings of the International Conference on Machine Learning*, pp. 7354–7363, Long Beach, CA, USA, 2019.

[125] D. Korkinof, T. Rijken, M. O'Neill, J. Yearsley, H. Harvey, and B. Glocker, "High-resolution mammogram synthesis using progressive generative adversarial networks," 2018.

[126] A. Khatami, A. Khosravi, C. P. Lim, and S. Nahavandi, "A Wavelet Deep Belief Network-Based Classifier for Medical Images," *Neural Information Processing*, in *Proceedings of the International Conference on Neural Information Processing*, pp. 467–474, Berlin, Germany, 2016.

[127] Y. Jia, E. Shelhamer, J. Donahue et al., "November. Caffe: convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM International Conference on Multimedia*, pp. 675–678, Orlando, FL, USA, November 2014.

[128] M. Abadi, P. Barham, J. Chen et al., "Tensorflow: a system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, pp. 265–283, 2016.

[129] D. Maclaurin, D. Duvenaud, and R. P. Adams, "Autograd: effortless gradients in numpy," *ICML 2015 AutoML workshop*, vol. 238, no. 5, 2015.

[130] R. Collobert, K. Kavukcuoglu, and C. Farabet, "Torch7: a matlab-like environment for machine learning," *BigLearn, NIPS Workshop*, vol. 34, 2011.

[131] Chollet, F., 2018. Introduction to keras. March 9th.

[132] A. Zagalsky, J. Feliciano, M. A. Storey, Y. Zhao, and W. Wang, "February. The emergence of github as a collaborative platform for education," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work Social Computing*, pp. 1906–1917, Vancouver, Canada, 2015.

[133] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multiscale Convolutional Neural Networks for Lung Nodule Classification," *Lecture Notes in Computer Science*, vol. 34, pp. 588–599, 2015.

[134] J. Kawahara, A. BenTaieb, and G. Hamarneh, "April. Deep features to classify skin lesions," *Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, vol. 32, pp. 1397–1400, 2016.

[135] M. Ghafoorian, N. Karssemeijer, T. Heskes et al., "Deep multi-scale location-aware 3D convolutional neural networks for automated detection of lacunes of presumed vascular origin," *NeuroImage: Clinical*, vol. 14, pp. 391–399, 2017.

[136] M. Simonovsky and N. Komodakis, "Onionnet: sharing features in cascaded deep classifiers," 2016.

[137] A. M. Rather, A. Agarwal, and V. N. Sastry, "Recurrent neural network and a hybrid model for prediction of stock returns," *Expert Systems with Applications*, vol. 42, no. 6, pp. 3234–3241, 2015.

[138] R. S. Lee, F. Gimenez, A. Hoogi, and D. Rubin, "Curated breast imaging subset of DDSM," vol. 8, The Cancer Imaging Archive, 2016.

[139] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.

[140] L. Fei-Fei, R. Fergus, and P. Perona, "June. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories," in *Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop*, p. 178, IEEE, Washington, D.C., USA, 2004.

[141] Y. LeCun, F. J. Huang, and L. Bottou, "June. Learning methods for generic object recognition with invariance to pose and lighting," *2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004*, vol. 2, pp. II–104, 2004.

[142] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," 2009.

[143] L. Li Deng, "The mnist database of handwritten digit images for machine learning research [best of the web]," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141-142, 2012.

[144] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: a database and web-based tool for image annotation," *International Journal of Computer Vision*, vol. 77, no. 1–3, pp. 157–173, 2008.

[145] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "June. Imagenet: a large-scale hierarchical image database," in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, Long Beach, CA, USA, June 2009.

[146] E. G. Adagbasa, S. A. Adelabu, and T. W. Okello, "Application of deep learning with stratified K-fold for vegetation species discrimination in a protected mountainous region using Sentinel-2 image," *Geocarto International*, vol. 43, pp. 1–21, 2019.

[147] X. Sun, A. Gossmann, Y. Wang, and B. Bischt, "December. Variational resampling based assessment of deep neural networks under distribution shift," in *Proceedings of the 2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1344–1353, IEEE, Xiamen, China, December 2019.

[148] B. Abdualgalil and S. Abraham, "February. Applications of machine learning algorithms and performance comparison: a review," in *Proceedings of the 2020 International Conference on Emerging Trends in Information Technology and Engineering (Ic-ETITE)*, pp. 1–6, IEEE, Vellore, India, 2020.