DIAGNOSTIC ACCURACY OF CLINICAL DECISION SUPPORT SYSTEMS (CDSSs) ORADIII AND ORAD DDX IN COMPARISON TO HISTOPATHOLOGICAL DIAGNOSIS OF JAW LESIONS

THESIS

Submitted to

KATHMANDU UNIVERSITY

SCHOOL OF MEDICAL SCIENCES

In partial fulfilment of the requirements for the degree of

MASTER IN MEDICAL RESEARCH

DR. Harleen Bali

2025

In collaboration with the Nepal Health Research Council

This work is dedicated to my dear students,

ma pa

and

Jagteshwar-the light of my life

KATHMANDU UNIVERSITY
SCHOOL OF MEDICAL SCIENCES
Collaborative Program of Kathmandu University and Dhulikhel Hospital

DECLARATION BY THE CANDIDATE

I hereby declare that this dissertation entitled **"Diagnostic Accuracy of Clinical Decision Support Systems (CDSSs) ORADIII and ORAD DDx in comparison to histopathological diagnosis of jaw lesions"** is a genuine and authentic research work carried by me under the support and guidance of **Prof. Dr. Dashrath Kafle**, Clinical Head, Dental Program and HOD in Department of Orthodontics, Kathmandu University School of Medical Sciences, Dhulikhel, Kavre.

.....

Dr. Harleen Bali

KU Regd. No:

Date:

KATHMANDU UNIVERSITY
SCHOOL OF MEDICAL SCIENCES
Collaborative Program of Kathmandu University and Dhulikhel Hospital

CERTIFICATE BY THE PRECEPTOR

This is to certify that the thesis entitled **"Diagnostic Accuracy of Clinical Decision Support Systems (CDSSs) ORADIII and ORAD DDx in comparison to histopathological diagnosis of jaw lesions"** is an authentic research work done by **Dr. Harleen Bali** under my guidance as a partial requirement as per Kathmandu University norms for the degree of **Master of Medical Research** in collaboration with NHRC **.**

……..……

Prof. Dr. Dashrath Kafle

Director, Dental Program,

Head of Department, Orthodontics and Dentofacial Orthopedics

KUSMS, Dhulikhel, Kavre

Date:

Government of Nepal

Nepal Health Research Council (NHRC)

Estd: 1991

CERTIFICATE

This is to certify that the thesis entitled **"Diagnostic Accuracy of Clinical Decision Support Systems (CDSSs) ORADIII and ORAD DDx in comparison to histopathological diagnosis of jaw lesions"** is an authentic research work done by **Dr. Harleen Bali** under the guidance of **Prof. Dr. Dashrath Kafle**, in partial fulfillment of the requirement for the degree of "**Masters in Medical Research**" granted by Kathmandu University in collaboration with Nepal Health Research Council (NHRC).

……..

Prof. Dr. Gehanath Baral

Program Coordinator,

Masters in Medical Research

Date:

ACKNOWLEDGEMENT

I wish to express my heartfelt gratitude to Our Dean, KUSMS, Prof. Dr. Manoj Humagain Sir and Our program Co-ordinator Prof. Dr. Gehanath Baral sir for their unwavering efforts in introducing Masters in Medical Research course and toiling hard towards making it organized and fruitful. Baral sir, you are one of the most authentic and supportive coordinator any student can ask for. Thank you.

My teachers/facilitators from NHRC and esteemed colleges of Nepal who shared not only their knowledge but their experience. Thank you

Dear Prof. Dr. Dashrath Kafle, Sir, though you have had and will have many students under your wings, I hope you recall my name among the ones that fill you with pride. You were instrumental in showing me my true strengths. I will always commend you for your out of box thinking, which is astonishing, and be in awe of the way your visionary mind works.

My modest appreciation to Dr. Sagar Adhikari, my partner in crime, for his patience and understanding, Dr. Nitesh Kumar Chaurasia, entire Oral and Maxillofacial Surgery Department and Dr. Pratibha Poudel, Dr. Bhoj Raj Adhikari, Oral and Maxillofacial Pathology Department for being ever ready to help out. Gita, Sachita, Khushbu, Manisha and Anil Sir for their compassion, encouragement and friendship. Dear colleagues and friends, thank you.

This acknowledgement would be incomplete without the mention of Prof. Dr. Chandan Upadhyaya, the then Director, Dental Program, who allowed us leave every time the classes were offline.

I also am grateful to OMR Dental Officer Krishna and intern Garima for being ready for more responsibilities if and when needed.

Ravi and Srijana, technicians, the unsung heroes of Oral Medicine and Radiology Department who work tirelessly and it is the result of their efforts that the archives of radiology is always available for research work.

I have been lucky to have been part of this encouraging and lively MMR batch 2024-2025. My batchmates.. Dear comrades you made all the difference. Made learning fun.

I am eternally grateful to all the subjects whose records made this study.

Thank you all for being a part of this journey.

To my higher self – hope we can be at peace again. Aham Brahmasmi

ਹਰਲੀਨ (**Harleen**)

LIST OF ACRONYMS

AI – Artificial intelligence

CDSS – Clinical decision support systems

OPG – Orthopantomogram

TP – True positive

FP – false positive

TN – True negative

FN – False negative

PPV- Positive Predictive Value

NNV- Negative Predictive Value

Se-Sensitivity

Sp-specificity

LR- Likelihood Ratio

CT – Computed tomography

CBCT – Cone beam computed tomography

MRI – Magnetic resonance imaging

USG – Ultrasonography

ABSTRACT

Background: With the coming age, integration of Artificial Intelligence is seen in almost all aspects of life, even medical field especially the field of radiology. Clinical Decision Support Systems (CDSSs) like ORADIII and ORAD DDx are available to help diagnose oral intra bony lesion. However, their diagnostic validity remains to be fully established and limitations need to be explored, especially when compared to gold standard of Diagnoses i.e. histopathological diagnosis.

Objectives: To evaluate and compare the diagnostic performance of two CDSS tools—ORADIII and ORAD DDx—against histopathological diagnosis in identifying intra-bony jaw lesions using orthopantomograms (OPGs).

Materials and Method: A cross-sectional diagnostic accuracy study was conducted on a sample comprising both lesion and non-lesion cases based on radiographic evaluation. Diagnostic outputs from ORADIII and ORAD DDx were compared with histopathology. Key performance indicators—including sensitivity, specificity, accuracy, F1 score, Positive Predictive Value (PPV), Negative Predictive Value (NPV), and likelihood ratios (LR^+ and LR^-)—were calculated for both systems. Concordance, partial concordance, and discordance with histopathological diagnosis were also assessed.

Results: Among the 350 samples evaluated, including 175 lesion-positive and 175 non-lesion cases, ORAD DDx demonstrated superior diagnostic performance compared to ORADIII. The sensitivity, specificity, accuracy, and F1 score for ORADIII were 64.57%, 60.00%, 62.29%, and 0.6314, respectively. In contrast, ORAD DDx achieved sensitivity, specificity, accuracy, and F1 score of 70.29%, 65.71%, 68.57%, and 0.6869 respectively. The positive predictive value (PPV) and negative predictive value (NPV)

for ORADIII were 61.75% and 62.87%, while for ORAD DDx, these were 67.21% and 68.86%, respectively. The positive likelihood ratio (LR+) and negative likelihood ratio (LR−) were 1.614 and 0.5905 for ORADIII, compared to 2.050 and 0.4513 for ORAD DDx.

Conclusion: In this study, ORADIII and ORAD DDx demonstrated moderate diagnostic performance when compared to the gold standard of histopathological diagnosis. ORAD DDx showed slightly higher sensitivity, specificity, and diagnostic accuracy than ORADIII, with greater concordance in identifying intra-bony jaw lesions. However, both systems provide radiographic-level diagnoses and do not replace histopathological evaluation. Their utility lies in supporting clinical decision-making, and while they show promise, further refinement and validation are needed before clinical integration as reliable diagnostic tools.

Keywords: *artificial intelligence; clinical decision support systems; jaw diseases/diagnosis; diagnostic imaging.*

TABLE OF CONTENTS

INTRODUCTION

The clinician's preliminary diagnosis and the pathologist's conclusive histopathological diagnosis, which is considered the gold standard, are not always the same. Due to time constraints and the high volume of patients, clinicians often rely on initial impressions or memorable experiences, which can potentially lead to diagnostic inaccuracies.

Artificial Intelligence (AI), with its data-driven approach, offers a more objective analysis free from personal biases. AI, which simulates human intelligence through machine-based algorithms, is increasingly integrated into everyday life and healthcare.[1]

Clinical Decision Support Systems (CDSSs), combine demographic, clinical, and radiological data, to generate differential diagnoses, providing clinicians with additional decision support. This, in turn, increases the efficiency and accuracy of clinical diagnoses.[2–5] CDSSs have demonstrated high accuracy rates in internal medicine, especially when diagnosing common chief complaints.[6]

A healthcare information technology company, in Nepal has developed a system to deliver services to patients, hospitals, and doctors.[7] A US-based Nepali AI scientist, Dr. Sameer Maskey, an adjunct associate professor at Columbia University and promoter of Fusemachines Inc., developed the first software robot in Nepal which is being used in e-commerce, trekking, airlines, hotels, hospitals, telecommunications, and government projects.[8]

Bayesian Belief Networks (BBNs) are another tool applied in both medical and dental fields to enhance diagnostic accuracy.[9–13] An example is the Oral Radiographic Differential Diagnosis (ORADIII) system developed by the University of California,

Los Angeles (UCLA) in the 1990s, which employs the Bayesian approach to diagnose intra-bony lesions of the jaw.[12] Using AI in decision support for diagnosing bony jaw lesions (including cysts and tumors) lies in its potential to aid in improving clinician's diagnostic accuracy, efficiency, and consistency. AI can further help with complex diagnoses, reduce diagnostic errors, provide support in high-volume or remote settings, as well as work educational tool for students.

Application of Proven Systems: Systems like the Dxplain,[14] Isabel[15] and Doknosis[16] are used by physicians as aids in diagnosis. Therefore, there is a need to study the diagnostic accuracy of the CDSS available in the field of Dentistry.

Dr. White, UCLA, developed a system named "ORAD" (https://www.orad.org/) in 1995, based on probabilistic/Bayesian calculations using conditional probabilities—the odds that a particular pathology would have a specific imaging feature and the prevalence of the pathology in the target population.[17]

This system has been upgraded to ORADIII and it provides differential diagnosis of identify intra-bony lesions. A user has to put the patient's clinical and radiographic features into the system, and a list of differential diagnoses is generated by it.[4] It is very useful as an adjunct for the general dentist in diagnosing oral bony pathologies.[18]

ORAD DDx (https://www.dentistry.nus.edu.sg/orad-ddx/) was developed at the National University of Singapore using information about lesion features from a textbook, "Oral Radiology: Principles and Interpretation".[19] It is a logical/deductive system based on an analytic/ System 2 approach that produces a list of possible differentials based on inputs/ filters (radiographic features) that users select.[17]

While CDSSs show potential in improving diagnostic and treatment decisions in radiology, oral radiology, and dentistry, direct comparison studies with histopathological outcomes remain scarce. Most available research evaluates CDSSs for decision support or primary care enhancement rather than diagnostic confirmation against gold standards like histopathology, and none so far are available comparing ORAD and ORAD DDx with histopathology diagnosis.

Therefore, this study aims to evaluate how accurately these CDSS tools (ORADIII and ORAD DDx) can match histopathological diagnoses for intra-bony jaw lesions of patients visiting Kathmandu University School of Medical Sciences, Dhulikhel Hospital, Kavre, Nepal.

REVIEW OF LITERATURE

1. Introduction to Jaw Lesions

- Classification
- Prevalence
- Clinical and radiological challenges in diagnosis
- Importance of accurate diagnosis for treatment planning and prognosis

2. Histopathology: The Gold Standard

- Role of histopathological examination
- Limitations
- Need for adjunctive diagnostic tools

3. Clinical Decision Support Systems (CDSSs)

- Definition and overview of CDSSs
- History and development of CDSSs
- Types of CDSSs
- Benefits

4. ORAD III and ORAD DDx Systems

- Development background

5. Description of ORAD III:

- Input parameters
- Output format

1. Description of ORAD DDx:

- Enhanced diagnostic features.
- Differences from ORADIII

2. Previous studies and validations.

3. Diagnostic Accuracy and Evaluation Metrics

- Definitions:

Sensitivity, specificity, Positive predictive value (PPV), Negative predictive value (NPV), Accuracy.

4. Comparative Studies on CDSSs vs. Histopathology

- Literature comparing CDSSs with histopathological outcomes
- CDSS in radiology
- CDSS in Dentistry

5. Gaps in existing literature regarding jaw lesions and dental applications

1. JAW LESIONS

The majority of lesions of the jaws originate as a sequel of pulpo-periapical pathologies of odontogenic origin. While others may be a result of pathologies related to the remnants of embryonic structures involved in jaw or tooth development. [20]

CLASSIFICATION

There are many ways to classify lesions of the jaws. The following two are the most followed by Oral Radiologists, Oral Pathologists, and Dentists in general:

I. 2022 WHO classification of odontogenic tumors and cysts of jaws.[21]

II. Differential diagnosis of Bony Lesions. [22]

Table I: 2022 WHO classification of odontogenic tumors and cysts of the jaws.[21]

ODONTOGENIC TUMORS
Benign epithelial odontogenic tumors
Adenomatoid odontogenic tumor
Squamous odontogenic tumor
Calcifying epithelial odontogenic tumor
Ameloblastoma, unicystic
Ameloblastoma, extraosseous/peripheral
Ameloblastoma, conventional

Adenoid ameloblastoma

Metastasizing ameloblastoma

Benign mixed epithelial & mesenchymal odontogenic tumours

Odontoma

Primordial odontogenic tumor

Ameloblastic fibroma

Dentinogenic ghost cell tumor

Benign mesenchymal odontogenic tumors

Odontogenic fibroma

Cementoblastoma

Cemento-ossifying fibroma

Odontogenic myxoma

Malignant odontogenic tumors

Sclerosing odontogenic carcinoma

Ameloblastic carcinoma

Clear cell odontogenic carcinoma

Ghost cell odontogenic carcinoma

Primary intraosseous carcinoma, NOS

Odontogenic carcinosarcoma

Odontogenic sarcomas

CYSTS OF THE JAWS

Radicular cyst

Inflammatory collateral cysts

Surgical ciliated cyst

Nasopalatine duct cyst

Gingival cysts

Dentigerous cyst

Orthokeratinised odontogenic cyst

Lateral periodontal cyst and botryoid odontogenic cyst

Calcifying odontogenic cyst

Glandular odontogenic cyst

Odontogenic keratocyst

…..

Non-Odontogenic Lesions

Giant cell granuloma (CGCG)

Fibrous dysplasia

Aneurysmal bone cyst (ABC)

Traumatic bone cyst (simple bone cyst)

II. Differential diagnosis of Bony Lesions.[22]

A. Radiolucencies of the Jaws

Anatomic Radiolucencies,

Periapical Radiolucencies,

Pericoronal Radiolucencies,

Interradicular Radiolucencies,

Solitary Cyst like Radiolucencies Not Necessarily Contacting Teeth,

Multilocular Radiolucencies,

Solitary Radiolucencies with Ragged and Poorly Defined Borders,

Multiple Separate, Well-Defined Radiolucencies,

Generalized Rarefactions of the Jawbones,

B. Radiolucent Lesions with Radiopaque Foci or Mixed Radiolucent-Radiopaque Lesions

Mixed Radiolucent-Radiopaque Lesions Associated with Teeth,

Mixed Radiolucent-Radiopaque Lesions Not Necessarily Contacting Teeth,

C. Radiopacities of the Jawbones

Anatomic Radiopacities of the Jaws,

Periapical Radiopacities,

Solitary Radiopacities Not Necessarily Contacting Teeth,

Multiple Separate Radiopacities,

Generalized Radiopacities.

The former classification is based on histopathology, whereas the latter is a radioimaging-based differential diagnosis of lesions in the jaws.

PREVALENCE

An article published in 2022, Clinico-pathological study of odontogenic cysts and tumours conducted at a tertiary care dental hospital of Nepal, reported that in a total of 163 biopsies, the majority, 73.62% cases were of odontogenic cysts, and the rest, 26.38% cases were odontogenic tumors. They found Radicular cyst and conventional ameloblastoma to be the commonest cysts and tumors.[23]

CLINICAL AND RADIOLOGICAL CHALLENGES IN DIAGNOSIS

Other than clinical, three approaches exist to developing a differential diagnosis or confirming a diagnosis. The radiological, pathologic, or surgical diagnosis. Imaging features of a central lesion, based on principles of interpretation, are used for radio-diagnosis and its differential, by Oral and maxillofacial (OMF) radiologists, who

comprise a subspecialty of dentistry. It requires systematic analysis and categorization of lesions based on various features. [19]

Step 1: Localize Abnormality

Step 2: Assess Periphery and Shape

Step 3: Analyze Internal Structure

Step 4: Analyze Effects of Lesion on Surrounding Structures

Step 5: Formulate Interpretation

IMPORTANCE OF ACCURATE DIAGNOSIS FOR TREATMENT PLANNING AND PROGNOSIS.

Imaging features of multiple lesions within the jaws resemble each other and are not always unique and distinct from their pathology. This somewhat puts the radio-diagnosis under the radar.[24]

Pathology gives a diagnosis based on the tissue of origin and changes seen at that level, thereby making biopsy a more reliable method and the gold standard for the diagnosis of many jaw lesions.[25]

2. HISTOPATHOLOGY: THE GOLD STANDARD

ROLE OF HISTOPATHOLOGICAL EXAMINATION:

a. Diagnosis of lesions with variant imaging features as per its stage:

Some lesions, for example, Eosinophilic Granuloma, show imaging features based on the stage of the lesion. Starting as a radiolucent lesion, with ill-defined borders to well-defined, and in the final stage, it presents with sclerotic borders. This makes diagnosis based on radiographic features as not very reliable.[26]

b. Diagnosis of lesions with resembling imaging features: Dentigerous cyst and Odontogenic Keratocyst, both present radio graphically as unilocular, well-defined radiolucency containing a tooth. This description is not enough for the operating surgeon since the pathology and management of both entities are different. The former being benign while the latter is known for its high recurrence rate.[20]

c. Differentiate benign from malignant lesions: Histopathology picture helps to differentiate benign from malignant

d. Aid in Treatment planning: With the accurate and confirmed diagnosis, an appropriate treatment plan can be formulated.

LIMITATIONS:

a. Invasive: Biopsy incisive or excisional involves cutting into the tissue, at times from multiple sites.

b. Time-consuming: starting with blood parameters to biopsy to lab procedure, and finally reading the prepared slide involves various steps, procedures, and equipment, making it lengthy.

c. Expensive: Since these procedures cost a lot, therefore, though histopathology is considered the gold standard in the diagnosis of jaw lesions, there is a need to find

alternative methods of diagnosis that are less invasive, inexpensive, and less time-consuming.[27]

NEED FOR ADJUNCTIVE DIAGNOSTIC TOOLS:

Since biopsy is invasive, options for non-invasive methods are in demand and being explored extensively. Vital staining, use of autofluorescence, chemiluminescence, USG, imaging modalities, plain CT, CBCT, and MRI are some such techniques. With the evolution in the field of Artificial Intelligence (AI). The non-invasive methods with the help of AI can help reach a diagnosis without the need for biopsies is to be explored.

3. CLINICAL DECISION SUPPORT SYSTEMS (CDSSS)

With the advancement in the field of technology and the new era of artificial intelligence (AI). The world is changing to adapt to the use of AI in the field of medicine. Though AI is known to take over many fields, in the field of medicine, it can be used as a supportive or adjunct tool rather than a replacement of physicians and doctors.

Even then, AI can be a very good assisting tool to doctors in general, regardless of their specialty. With the availability of raw data for machine learning and support of analytics under the guidance of experienced doctors, AI can help in clinical decision making, treatment planning, and progress evaluation.[28]

A type of AI is clinical decision support systems (DSS) that intelligently filters knowledge and patient information to provide diagnoses and evidence support for clinical decisions. They are aimed at assisting the physician in decision-making rather than replacing him. [18]

HISTORY AND DEVELOPMENT OF CDSSS.

Studies regarding the use of computers as a support system for professionals began as early as 1950. The first evidence for its use for medical purposes is seen in the paper 'Reasoning Foundations of Medical Diagnosis' published in the late 1950s by Ledly and Lusted. This opened a gateway for more exploration along the road.[29]

They reported that diseases and their manifestations can be linked using punch cards. The resistance to accept something not part of medical education and the inadequacies of knowledge of the field of computers acted as a hurdle to its acceptance.

F.T. de Dombal et al. developed Leeds abdominal pain system using Bayesian probability theory. The Pathfinder system for the diagnosis of lymph node pathology was also made on similar basics.[30]

MYCIN8, the first rule-based system, was developed in 1970. It further led to more systems based on a similar model.[31]

Hybrid systems now combine deductive rules and probabilistic reasoning in the same CDSS. Best known of the hybrid systems are the general medical consultation systems QMR11 (1985), DXplain12 (1986), and Iliad13 (1987).[30]

The 2010s saw a surge in AI and machine learning (ML) techniques, which improved the development of CDSSs. Examples include IBM Watson Health and Google's DeepMind, combinations of AI and ML, that are transforming healthcare decision-making.[29]

In the era of mobile and Telemedicine, CDSSs have risen beyond the culture of traditional one-to-one clinical consultations. Mobile health (mHealth) applications and

remote monitoring tools have incorporated CDSSs to help patients and medical practitioners outside the clinical setup, making it accessible and time-saving.[29]

CDSSs are being trained to include patients' preferred treatment options, making practitioners understand patients' needs and plan treatment accordingly. Ruland et al. observed that practitioners were able to provide better patient-centered treatment plans if patients' symptoms and preferences were taken into account.[30]

CDSS use in dental clinics has been recommended to be classified into either static or dynamic. Static systems are unable to upgrade in terms of new information, whereas dynamic systems can do so. Since dynamic systems have machine learning features, they can support a real-time, individualized plan by taking into account the profile of each patient.[18]

To understand this better, let us assume a patient reported to the clinic with the chief complaint of a toothache and filled out a questionnaire provided by the CDSS. The system will itself generate a treatment plan based on the information provided, which can include symptoms, dietary habits, fluoride exposure, and past dental history etc. This can help dentists include all relevant information, not overlook any option, and provide better individualized patient care.[32]

TYPES OF CDSSS:

CDSS, as artificial intelligence (AI) help support clinical decision making. The two main categories of AI uses in CDSS are usually noted:

a) Knowledge-based AI (also called rules-based expert systems) and

b) Data-driven AI.

Initial systems are the knowledge-based AI ones, which mimic human decision making by using rules laid by field experts in the medical field in software terms. Rules such as in case a patient reports symptoms A, use medication B. Thus, such logic can be easily traced to its origin and reassessed.[33]

The data-driven AI has come up in the recent decade. It uses machine learning algorithms to draw patterns from huge raw data. Training datasets containing data from patient records previously treated by practitioners are fed to the system as part of training it. The CDSS thus learns to recognize or track a pattern that fits best with good health care outcome or treatment plan. On entering a new case into the system, the system uses the learned pattern to recognize and diagnose.[33]

However, based on large data sets employed as a 'training set', the data-driven AI can predict subtle changes and catch minute details, but unlike knowledge-based AI, the decision given cannot be easily tweaked and evaluated. This makes their reliability and accountability questionable.[33]

BENEFITS OF AI

1. Time saving

2. Using all the available information logically to provide accurate diagnosis.

3 Procedures can be detailed, standardized and reproducible.[32]

4. Early recognition of certain diseases without overlooking any possibility

5. Clinical organization: Provides regular reminders, advises on cautions, keeps and maintains records. Improve work-flow.[34]

4. ORAD III and ORAD DDx Systems

DEVELOPMENT BACKGROUND:

Dr. White, UCLA, developed a system named "ORAD" (https:// www. orad. org/) in 1995, based on probabilistic/Bayesian calculations using conditional probabilities—the odds that a certain pathology would have a precise imaging feature and the prevalence of the pathology in their target population.[17]

This system has been upgraded to ORADIII and it generates differential diagnosis of recognize intra-bony lesions. A user has to put the patient's clinical and radiographic features into the software, and a series of differential diagnoses is reported by it.[4] It is very beneficial as an aide for the general dentist in diagnosing oral pathologies.[18]

According to White there is an extensive range of lesions that may have radiographic picture in the jaws. Often, these lesions are challenging to deduce because their radiographic features are not pathognomonic, but may resemble with various other lesions. Accordingly, the purpose of developing ORAD was to develop a program to support the dentist to frame differential diagnoses for radiographic lesions in the jaws. It should not be used as a substitute for clinical judgment: rather only as an assistance to the clinician in proposing conditions not formerly considered.[12]

Accordingly, this is to help the clinician think broadly and to consider a wider range of possibilities when evaluating radiographic lesions. Most human errors in differential diagnosis result from errors of omission.[12]

ORAD DDx (https:// www.dentistry.nus.edu.sg/orad-ddx/) is a product of the National University of Singapore. It uses information about lesion features from the textbook,

"Oral Radiology: Principles and Interpretation".[19] It is a logical/deductive system based on an analytic/System 2 approach that produces a list of probable differentials based on inputs/ filters (radiographic features) that users select.[17]

5. DESCRIPTION OF ORAD III:

Input parameters [12]

Includes 16 questions, with options provided, to be answered from the options provided, based on the patient's clinical and radiographic features.

Patient characteristics

- How old is patient?
- What is the race of the patiene?
- What is the patient's gender?
- Does the patient have pain or paraesthesia?

Location of lesion

- The lesion origin
- Where is the lesion?
- Where is the lesion located?
- Is the lesion odontogenic in origin?

Lesion growth

- How many lesions are there?
- How big is the lesion?
- Is there bony expansion?
- Is the lesion loculated?
- the lesion borders?
- the lesion contents?
- Is there root resorption?
- Is there tooth displacement or impaction?

Output format

List of probability-based differential diagnoses in order from most likely with probability percent to least likely based on the features fed to it.

6. DESCRIPTION OF ORAD DDX:

Enhanced diagnostic features:

ORAD DDx (https:// www.dentistry.nus.edu.sg/orad-ddx/) is a deductive system created on an analytic/System 2 method that produces a list of probable differentials based on inputs/ filters (radiographic features) that operators input and presents a forward reasoning framework to the users. Makers claim this causes improved diagnostic accuracy.[17]

Filter options are:

- Number
- Epicenter is within the neurovascular canal
- Mandibular lesions: Epicentre is located below the IAC
- Lesion is associated with a single tooth peripex
- Lesion is in a follicular relationship with an associated tooth.
- Internal density (Radiolucent/Mixed/Radiopaque)
- Border definition (Well-defined/Ill-defined)
- Border cortication (Yes/No)
- Encapsulation within soft tissue border or PDL space (Yes/No)

DIFFERENCES FROM ORAD III

Both programs were developed by institutes to provide differential diagnosis based on radiographic features of oral lesions. ORAD was developed by UCLA in the 1990s, and ORAD DDx is comparatively a more recent model.

Makers of the ORAD DDx report that ORAD is based on probabilistic/ Bayesian calculations and the prevalence of the pathology in the target population. However, mistakes of these calculations can arise from the lack of prevalence of data (for that region).[17]

ORAD now upgraded to ORADIII takes account of clinical as well as radiographic features, whereas ORAD DDx is based on radiographic features only and was developed mostly to educate undergraduates.

7. PREVIOUS STUDIES AND VALIDATIONS:

a. Development of ORAD at UCLA; 1995

Dr. Stuart C. White at UCLA developed the original Oral Radiographic Differential Diagnosis (ORAD) system in 1995. Bayesian system of calculating probabilities was developed to help oral radiologists and dentists arrive at a diagnosis. It was developed with the aim of supporting and not replacing the actual practitioner.[35]

One downside of the system is that the program could not consider lesions not entered into its knowledge bank. Practical limitations of memory space, lesion data availability, and computational speed led to the knowledge base being finite; to include only intrabony lesions described in dental radiology and oral pathology textbooks.[17]

b. Evaluation of ORAD's Diagnostic Validity; 2012

A.F. Simeos et al. in their study assessed ORAD's diagnostic accuracy in identifying jaw bone pathologies. The findings indicated that 67% of ORAD's radiographic diagnoses did not match histopathological diagnoses, suggesting that while ORAD could be a helpful assistance, it cannot replace expert clinical judgment.[36]

c. Case Series Using ORADIII; 2017

S.L. Brooks conducted a case series using ORADIII on five different bony jaw lesions. The study observed that though the software is useful but its accuracy heavily depends on the precision of the input data provided.[36]

d. Comparative Study with Specialists; 2018

Vicari AP et al. did a study comparing ORAD's diagnostic performance with that of dental specialists in interpreting panoramic radiographs. The study reported a sensitivity of 87.5% for ORAD and 93.75% for specialists, concluding that ORAD could serve as a supportive tool in cases where the presence of pathology is assured.[36]

e. ORAD DDx vs. Atlas in Dental Education; 2022

This study was done to validate ORAD DDx, assessing the efficiency of ORAD DDx in comparison to an Atlas-based approach in improving dental students' diagnostic accuracy. The results suggested that the Atlas group outperformed both the ORAD DDx and control groups in diagnostic accuracy and recall of radiographic features. However, students reported that both ORAD DDx and the Atlas augmented their confidence and reduced the mental effort required for coming up with the differential diagnoses.[17]

f. Comparing ORAD and Radiologists, 2024

A cross-sectional study compared the diagnostic accuracy of ORADIII software with that of maxillofacial radiologists in diagnosing benign jaw lesions. The study found that ORAD had a diagnostic accuracy of 50%, while radiologists achieved 68.4%. The difference was statistically insignificant ($p = 0.103$), suggesting that ORAD could be a useful adjunctive tool but should not replace expert evaluation.[36]

8. Diagnostic Accuracy and Evaluation Metrics

DEFINITIONS:

Sensitivity

The proportion of true positives (TP) tests out of all patients with a condition is termed as sensitivity. It is the ability of a test or instrument to yield a positive outcome for a subject who has that disease.[37] The equation for sensitivity is the following:

Sensitivity = (True Positives (A)) / (True Positives (A) + False Negatives (C))

Specificity

Specificity is defined as the percentage of true negatives (TN) out of all subjects who do not have a disease or condition. It is the ability of the test or instrument to obtain normal range or negative results for a person who does not have a disease.[37] The formula to determine specificity is the following:

Specificity = (True Negatives (D)) / (True Negatives (D) + False Positives (B))

Sensitivity and specificity are inversely related: as sensitivity increases, specificity tends to decrease, and vice versa. Sensitivity and specificity should always merit consideration together to provide a holistic picture of a diagnostic test.[37]

Positive Predictive Value (PPV) and Negative Predictive Value (NPV)

PPVs determine, out of all of the positive findings, how many are true positives;

NPVs determine, out of all of the negative findings, how many are true negatives.

As the value increases toward 100, it approaches a 'gold standard'.[37]

The formulas for PPV and NPV are below.

Positive Predictive Value = (True Positives (A)) / (True Positives (A) + False Positives (B))

Negative Predictive Value = (True Negatives (D)) / (True Negatives (D) + False Negatives(C))

Diagnostic effectiveness (accuracy)

Another global measure is diagnostic accuracy (effectiveness). It is the proportion of correctly classified subjects (TP+TN) among all subjects (TP + TN + FP+ FN).[37]

Diagnostic accuracy is affected by the disease prevalence. With the same sensitivity and specificity, the diagnostic accuracy of a particular test increases as the disease prevalence decreases.[38]

9. COMPARATIVE STUDIES ON CDSSS

CDSS WITH HISTOPATHOLOGY

Over the past decade, many systematic reviews have examined the performance of Clinical Decision Support Systems (CDSSs) with relation to histopathological diagnosis, particularly in the field of oncology and digital pathology.

a. AI in Digital Pathology: Diagnostic Accuracy Review

A comprehensive systematic review and meta-analysis by McGenity et al. (2023) evaluated 100 studies involving over 152,000 whole slide images across various disease types. The analysis reported a mean sensitivity of 96.3% and specificity of 93.3% for AI applications in digital pathology, using histopathological assessment and immunohistochemistry as reference standards. Despite these promising results, the study highlighted substantial heterogeneity in study designs and noted that all included studies had at least one area of high or unclear risk of bias.[39]

b. CDSSs in Oncology: Updated Systematic Review

An updated systematic review by Nafees et al. (2023) analyzed 43 studies on CDSSs used in oncology from 2016 to 2022. They found that 42 studies reported improvements in outcomes, with 34 demonstrating statistically significant results. The CDSS tools evaluated included computerized physician order entry systems, clinical practice guideline systems, patient-reported outcome tools, clinical pathway systems, and

prescriber alerts. The review noted that CDSSs can enhance guideline adherence, patient-centered care, and care delivery processes in oncology.[40]

c. CDSSs in Breast Cancer Treatment Decisions

A systematic review focused on breast cancer treatment decisions identified 17 studies evaluating CDSS tools. One notable study within this review assessed the impact of the OncoDoc CDSS. They noted that its use increased compliance with clinical practice guidelines from 61% to 77%. The study also reported that in cases where physicians' initial decisions were non-compliant, 62% changed their decisions to align with OncoDoc's recommendations.[41]

The mentioned reviews suggest that CDSSs, particularly those incorporating AI and digital pathology, show good diagnostic efficiency and can improve clinical decision-making.

However, the variability in study designs and the presence of biases question the reliability of CDSSs in clinical practice.

CDSS WITH RADIOLOGY

a. CDSS Application in Radiology:

Khalfallah et al. in 2023 published a systematic review discussing CDSS applications in diagnosis, monitoring, prediction, and recommendations across healthcare, including radiology. But it does not provide a direct comparison with histopathological results.[42]

CDSS Application in Dentistry:

a. CDSS for Dental Treatment of Fractured Teeth:

Zainuddin et al. (2013) developed a CDSS aimed at supporting treatment planning for fractured teeth; however, no histopathological validation was conducted.[43]

While CDSSs show potential in improving diagnostic and treatment decisions in radiology, oral radiology, and dentistry, direct comparison studies with histopathological outcomes remain scarce. Most available research evaluates CDSSs for decision support or primary care enhancement rather than diagnostic confirmation against gold standards like histopathology.

10. GAPS IN EXISTING LITERATURE ON JAW LESIONS AND DENTAL APPLICATIONS OF CDSSS

a. Lack of Direct Validation with Histopathology

o While Clinical Decision Support Systems (CDSSs) are emerging in dental radiology, very few studies compare CDSS diagnoses of jaw lesions directly with histopathological gold standards.

o Most current systems assist in preliminary diagnosis but lack validation through biopsy or histopathological reports, which limits their clinical reliability.

Nafees et al. (2023): Noted in oncology CDSS review that head and neck tumors have less robust decision support systems compared to other cancers.[40]

b. Small Sample Sizes and Lack of Multicenter Studies

• Many available studies involve small, single-center datasets, reducing the generalizability of findings.

- There is a need for multicenter, multi-population datasets to validate the diagnostic accuracy across diverse clinical settings.

c. Absence of Prospective Clinical Trials

- To date, most evaluations of CDSSs for dental applications are retrospective.

- There is a lack of prospective clinical trials that assess how CDSS usage influences real-time decision-making and treatment outcomes in dental and maxillofacial practice.

Ben Khalfallah et al. (2023), reviewed CDSSs in healthcare, identified that dental and maxillofacial applications are "emerging" but "understudied". [42]

d. Underrepresentation of Oral Radiologists in System Development

- Most CDSSs are developed by engineers or computer scientists, with limited involvement of oral radiologists during algorithm training and validation.

- This can lead to systems that miss the subtlety of clinical-radiological correlation that specialists are trained to recognize.

e. Inconsistent Reporting Standards

- Publications often lack standardized performance metrics like sensitivity, specificity, AUC (Area Under Curve), PPV, and NPV specific to jaw lesions.

- Reporting methods vary, making comparisons between studies difficult.

AI systems for histopathology are mainly developed for general cancer pathology, with minimal focus on oral/maxillofacial pathology.

OBJECTIVE

Primary Objectives:

- To assess the Diagnostic Accuracy of CDSS in jaw lesions in comparison to histopathology diagnosis.

Secondary Objectives:

- **Evaluate Diagnostic Accuracy**: To assess how accurately the CDSSs (ORADIII and ORAD DDx) predict or match the histopathological diagnosis of various intra-bony jaw lesions.
- **Compare Sensitivity and Specificity**: To compare the sensitivity (ability to correctly identify diseased cases) and specificity (ability to correctly identify non-diseased cases) of the CDSSs with the gold standard of histopathology.

RESEARCH HYPOTHESIS

Null Hypothesis: Clinical Decision Support Systems (CDSSs) ORADIII and ORAD DDx show diagnostic accuracy same as histopathological diagnosis of jaw lesions.

Alternate hypothesis:

Clinical Decision Support Systems (CDSSs) ORADIII and ORAD DDx show diagnostic accuracy different from histopathological diagnosis of jaw lesions.

MATERIALS AND METHODS

This cross-sectional, comparative, diagnostic accuracy study was conducted in the Department of Oral Medicine and Radiology, Kathmandu University School of Medical Sciences, Dhulikhel Hospital, Kavre, Nepal, from January 2025 to April 2025 after IRC approval. Patients with confirmed bony pathologies, i.e, Lesion Group (LG), via histopathology reports, involving either or both jaws; with demographic and panoramic radiographic (OPG) records from January 2019 to March 2025, retrospectively, and a Group of patients with radiographs (OPG) without any jaw lesions, i.e., Non Lesion Group (NLG) were included in the study. Non-probability convenience sampling was done.

Since ORADII has sensitivity= 66.67%, Specificity= 84.94%; Diagnostic Accuracy= 81.68%[44], sample size calculation was as:

$N_{LG} = [Z^2 \times (1 - \text{Sensitivity}) \times \text{Sensitivity}]/e^2$

Where: N = Sample size required.

- Z = Z-value (the number of standard deviations from the mean) corresponding to the desired confidence level i.e. for 95% confidence level, Z=1.96).
- Sensitivity = The estimated sensitivity of the diagnostic test i.e. 0.66 for 66.67%.[44]
- e = Desired margin of error for the estimate i.e. 0.07 for 7% margin of error.

Calculation:

$N_{LG} = [(1.96)^2 \times (1 - 0.66) \times 0.66]/(0.07)^2$

$N_{LG} = 175$

For samples without jaw lesions,i.e,, NLG:

$N_{NLG} = [Z^2 \times (1 - \text{Specificity}) \times \text{Specificity}]/e^2$

$N_{NLG} = [(1.96)^2 \times (1-0.8494) \times 0.8494] / (0.07)^2$

$N_{NLG} = 101$

Where:

- N_{NLG} is the number of non-pathology samples
- and the specificity of the test (as a proportion, so 84.94% = 0.8494)[44]

To make the two groups (LG and NLG) equal; 175 samples were taken in each group.

Ethical clearance was obtained from the Institutional Review Committee (IRC), Kathmandu University School of Medical Sciences (IRC-KUSMS Approval No: 303/24).

Inclusion criteria:

For the lesion group (LG): Biopsy reports of patients whose demographic data and panoramic radiographs are available were included. A biopsy report confirming a lesion diagnosis involving either jaw bone (maxilla or mandible or both) and panoramic radiographs of good quality were included.

Whereas for the Non-lesion group (NLG): Age and Gender matched patients with demographic and radiographic data without any pathology involving the jaws were taken. Panoramic radiographs of good quality were included.

Exclusion Criteria:

For Lesion Group: Biopsy reports of inadequate sample or unclear/overlapping histopathological diagnoses, incomplete patient records, and poor-quality panoramic radiographs

For Non Lesion Group: Incomplete patient records, and poor-quality panoramic radiographs

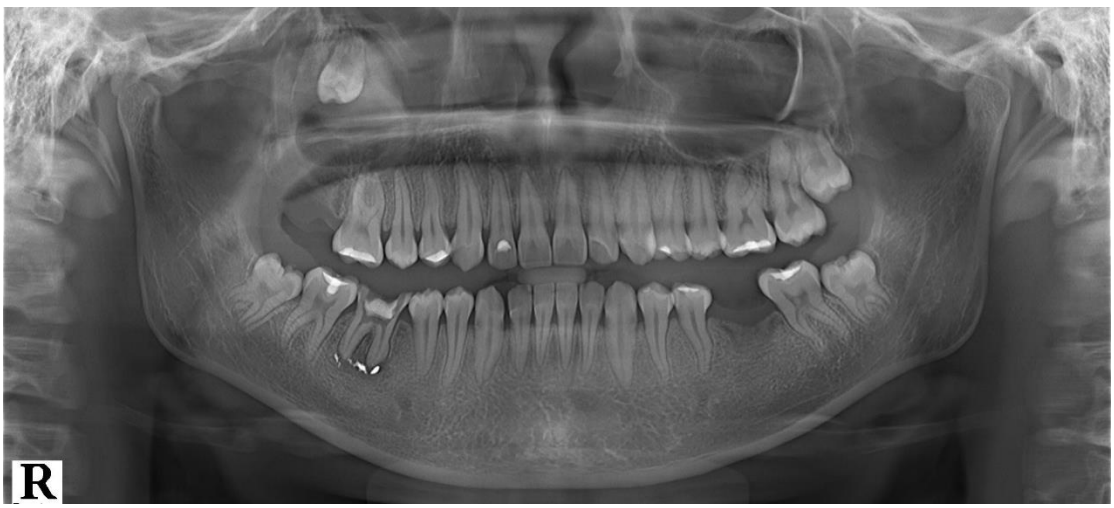
OPERATIONAL DEFINITIONS

Jaw lesions: Intra-bony lesions involving the maxilla or mandible or both present with some amount of osseous changes either in the form of rarefaction, remodeling or sclerosis on the radiograph. These changes are in relation to the degree of osseous rarefaction and remodeling differ among inflammatory, benign, and malignant lesions, and it is this feature, along with location of the lesion, which allows for its differentiation.[24]

Classification of jaw lesions was done as per WHO classification of odontogenic tumors and cysts of jaws, and anatomic variations was done as per the book Differential Diagnosis of Oral and Maxillofacial Lesions.[21, 22]

Othropantomogram/Panoramic Radiograph: An orthopantomogram (OPG) is a common radiograph used to identify the hard tissues of the maxilla, mandible and surrounding skeletal structures. Gross changes in mineralization of the dental structures, and changes in ossification of the underlying mandible and maxilla can aid in identification of inflammatory as well as diseases of developmental origin.[45]

Figure 1: Orthopantomogram taken using Dentium Rainbow CBCT machine.



CDSS: Clinical Decision Support System (CDSS) ia a type of Artificial Intelligence is a that intelligently filters knowledge and patient information to provide diagnosis and

evidence support for clinical decisions. They are aimed at assisting the physician in decision making rather than his replacement.[18]

ORADIII: ORAD, or oral radiographic differential diagnosis, is a pathology-related CDSS that was first developed by S.C. White in 1989. It is a computer software designed to assess the clinical and radiographic characteristics of patients who have intra-bony lesions in order to help identify those patients.[36]

This software can be accessed free of cost by visiting the site www.orad.org

Input paramenters: Includes 16 questions, with options provided, to be answered from the options provided, based on patient's clinical and radiographic features.

Patient characteristics

- How old is patient?
- What is the race of the patient?
- What is the patient's gender?
- Does the patient have pain or paraesthesia?

Location of lesion

- he lesion origin
- Where is the lesion?
- Where is the lesion located?
- Is the lesion odontogenic in origin?

Lesion growth

- How many lesions are there?
- How big is the lesion?
- Is there bony expansion?
- Is the lesion loculated?
- the lesion borders?
- the lesion contents?
- Is there root resorption?
- Is there tooth displacement or impaction?

Figure 2a: Opening page of the ORADIII site

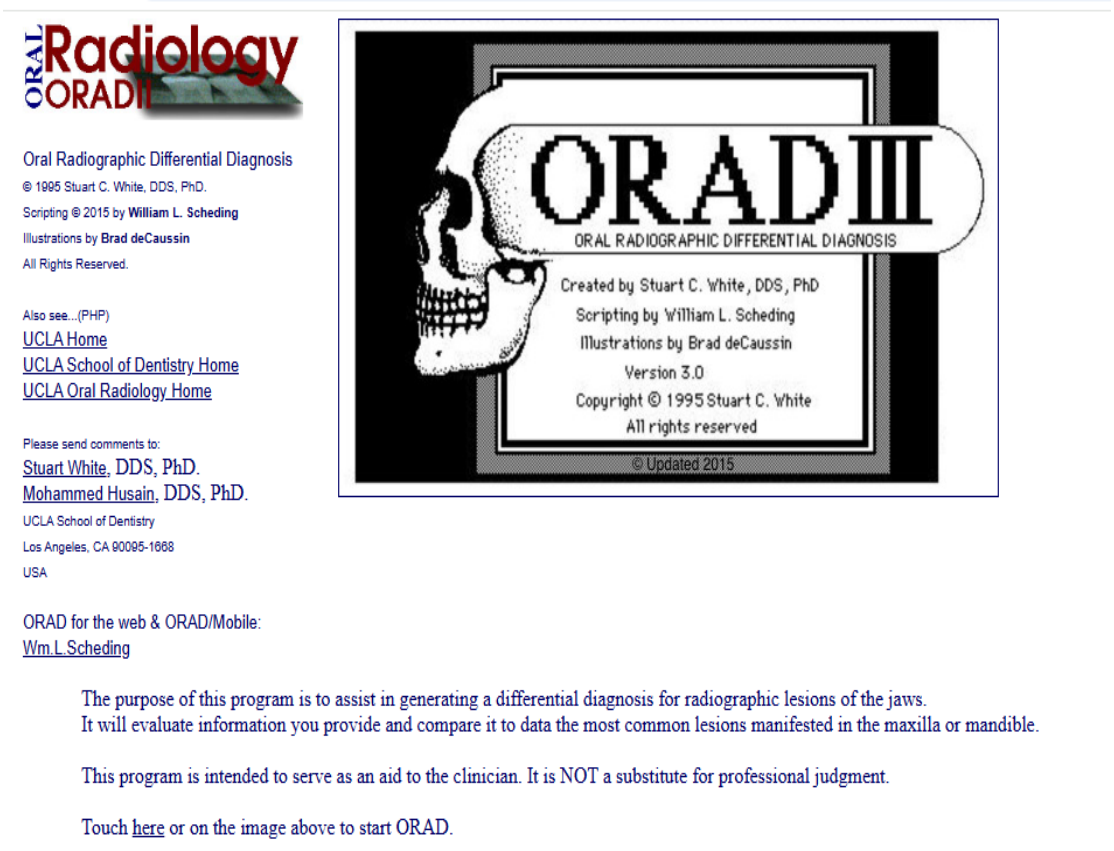


Figure 2b: ORAIII input page

Patient Characteristics

CLINICAL FEATURES

What is the sex of your patient? Male
What is the race of your patient? Nonblack
What is the age of your patient? 26 - 50
Does your patient have pain or paresthesia? No pain

RADIOGRAPHIC FEATURES

Location

Which jaw contains the lesion? Mandible only
The lesion center is in what region ? Molar region
The relationship of the lesion to teeth is: Not tooth associated
Please estimate the number of lesions: One
What is the maximum size of the lesion? Less than 2 cm
Where is the origin of the lesion? Central

Periphery

The borders of the lesion are: Corticated
The loculation of the lesion is: Unilocular

Internal Structure

The contents of the lesions are: Radiolucent
Does the lesion contain one or more teeth ? No

Effects on Surrounding Structures

Does the lesion expand the bony cortex? No
Does the lesion cause root resorption ? No
Does the lesion cause tooth displacement or impaction? No

Shall we consider prevalence ? Yes

Touch Differential when finished to formulate a radiographic differential.

Navigation: Home | Introduction | Patient Characteristics | Differential | Lesions

ORAD DDx: ORAD DDx (https://www.dentistry.nus.edu.sg/orad-ddx/) was developed at the National University of Singapore using information about lesion features from a textbook, "Oral Radiology: Principles and Interpretation". It is a logical/deductive system based on an analytic/System 2 approach that produces a list of possible differentials based on inputs/ filters (radiographic features) that users select. [17]

This software can be accessed free of cost by visiting the site.

Input parameters:

- Number
- Epicenter is within the neurovascular canal
- Mandibular lesions: Epicentre is located below the IAC
- Lesion is associated with a single tooth peripex
- Lesion is in a follicular relationship with an associated tooth.
- Internal density (Radiolucent/Mixed/Radiopaque)
- Border definition (Well-defined/Ill-defined)
- Border cortication (Yes/No)
- Encapsulation within soft tissue border or PDL space (Yes/No)

Figure 3a: ORAD DDx Input page

Figure 3a: ORAD DDx Input page continued

METHODS (PROCEDURE)

Histopathological reports were collected from an electronic database (MIDAS) and paper archives of the Oral Pathology Department, KUSMS, Dhulikhel Hospital. The diagnosis was according to the 2022 WHO classification or their alias based on previous classifications (2005 and 2017).[21,46] The OPD number of these cases would be used to find the demographic and panoramic (OPG) radiographs from the radiology archives of Department of Oral Medicine and Radiology, KUSMS, Dhulikhel Hospital. The radiographs were taken using two machines.

1. Dentium Rainbow CBCT machine. All Panoramic images were taken at 79 kVp, 10 mA, and scan time 19.0 sec, in standard Mode.

2. Planmeca 2D imaging machine. All Panoramic images were taken at 70 kVp, 14 mA, and scan time 17.0 sec, in standard Mode.

The cases meeting the inclusion and exclusion criteria were included in the study. The radiographs were anonymized. The demographic and radiographic data were then entered into the ORADIII (http://www.orad.org/cgi-bin/orad/index.pl)[12] and ORAD DDX (https://www.dentistry.nus.edu.sg/orad-ddx/)[17] systems by a Co-investigator, with seven years of experience in the field of Oral and maxillofacial radiology, who was blinded to the said groups and histopathological diagnosis of the same. Prevalence consideration in ORADIII and common in ORAD DDx was marked when searching the differential.

Clinicopathological concordance was noted as follows: Definition of different concordance categories:[44]

1) Concordance was when the first provisional/first diagnosis matched the definitive/histopathological diagnosis.

2) Partial concordance was when multiple diagnoses were given, of which included the correct diagnosis within first three on the list but not listed as the first/provisional diagnosis.

3) Discordant was defined as:

a. An incorrect first/provisional diagnosis

b. A widely termed clinical provisional diagnosis (e.g., "Cyst, Tumor" given as provisional diagnosis for an OKC case)

c. When differential diagnoses did not contain the histo-pathological diagnosis or in normal cases contained some pathological diagnosis.

The data was entered into MS EXCEL software (version 2019, Microsoft®, USA).

Statistical Methods

Concordance was compared between

1. ORADIII and histopathology diagnosis,
2. ORAD DDx and histopathology diagnosis

STATISTICAL ANALYSIS

The statistical analysis was performed using SPSS (Statistical Package for Social Sciences) Software Version 25 (IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.).

Intra-observer variability:

Same blinded observer re-entered the values for 30 samples after one month gap before the beginning of the research. The kappa value was calculated to assess intra- -observer reliability.

The intra-examiner reliability of the sample calculated using the Kappa test is given in Table 1. The test was carried out on a blinded data collected 1 months apart, for five radiographic features. All the measurements showed almost perfect to perfect intra-observer agreement.

Table 1: Intra-examiner reliability of the variables studied

Features	Kappa value
Location	1.00
Size	.969
Border	.810
Internal structure	1.00
Root resorption	.902

Statistical Test: Kappa

Diagnostic accuracy was calculated for each software. The calculation for sensitivity, specificity, Accuracy, F1 Score, Positive Predictive Value (PPV)/Precision, Negative Predictive Value (NPV), and likelihood ratio (LR) was determined using the calculation formulae as below.

Sensitivity (Se) = TP/ TP + FN

Specificity (Sp) = TN /TN + FP

Accuracy = TP + TN /TP + TN + FP + FN

F1 score = 2 ∗ TP /2 ∗ TP + FP + FN

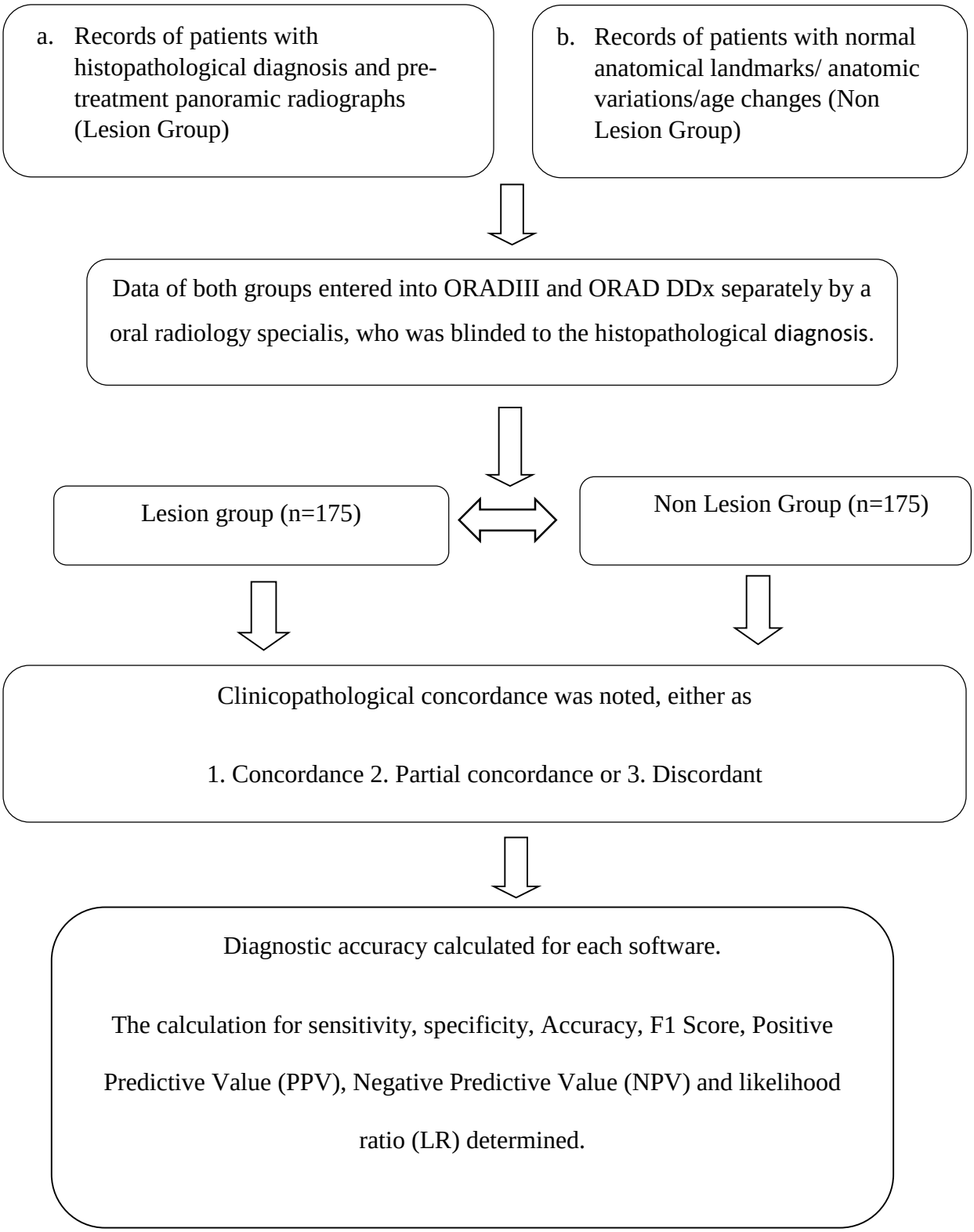
Positive Predictive Value (PPV) = TP/ TP + FP

Negative Predictive Value (NPV) = TN / TN + FN

Likelihood Ratio (LR+) = Se/(1-Sp)

(LR-) = Sp/(1-Se)

CONCEPTUAL FRAMEWORK



RESULT

A total of 452 histopathological diagnosis reports were retrieved from the archives of Oral pathology from January 2019 to March 2025. Of these, 175 biopsies with positive intra-bony jaw lesion diagnosis meeting the inclusion and exclusion criteria was admitted for the study in lesion group (LG) and 175 anatomic landmarks, normal anatomic variations and age changes were included in Non-lesion group (NLG).

Table 2: Comparison of ORAD III Diagnoses with Histopathological Findings in Lesion Detection

	ORADIII lesion positive	ORADIII lesion negative	Total
Histopathological lesion positive (LG)	113 *(TP)*	62 *(FN)*	175
NLG	70 *(FP)*	105 *(TN)*	175
Total	183	167	350

Cross-tabulation comparing ORADIII diagnoses with histopathological findings for lesion detection. Used to assess diagnostic performance metrics such as sensitivity, specificity, and predictive values.

Table 3: Comparison of ORAD DDx Diagnoses with Histopathological Findings in Lesion Detection

	ORAD DDx lesion positive	ORAD DDx lesion negative	Total
Histopathological lesion positive (LG)	123 *(TP)*	52 *(FN)*	175
NLG	60 *(FP)*	115 *(TN)*	175
Total	183	167	350

Cross-tabulation comparing ORAD DDx diagnoses with histopathological findings for lesion detection. Used to assess diagnostic performance metrics such as sensitivity, specificity, and predictive values.

Concordance between the histopathological diagnosis and ORADIII was in 90 (51.42%) cases, partial concordance in 23 (13.14%) and discordance in 62 (35%). Whereas, concordance between the histopathological diagnosis and ORAD DDx was in 116 (66.29%) cases, partial concordance in 7 (4%) and discordance in 52 (29.71%) among the present study sample.

Partial concordance was included in True positive, since final is from among provisional and differential diagnosis.

The sensitivity, specificity, accuracy and F1 score for ORADIII was 64.57%, 60%, 62.29% and 0.6314 respectively. Whereas, sensitivity, specificity, accuracy and F1 score for ORAD DDx was 70.29%, 65.71%, 68.57% and 0.6869 respectively among the study sample.

The Positive Predictive Value (PPV), Negative Predictive Value (NPV), Positive Likelihood Ratio (LR+) and Negative Likelihood Ratio (LR^-) for ORADIII were 61.75%, 62.87%, 1.614 and 0.5905 respectively. Whereas, The Positive Predictive Value (PPV), Negative Predictive Value (NPV), Positive Likelihood Ratio (LR+) and Negative Likelihood Ratio (LR^-) for ORAD DDx were 67.21%, 68.86%, 2.050 and 0.4513 respectively in the present study sample.

Table 4: Comparison of Diagnostic Performance Metrics Between ORAD III and ORAD DDx in Lesion Detection

Metric	**ORADIII**	**ORAD DDx**
Sensitivity	64.57%	70.29%
Specificity	60.00%	65.71%
Accuracy	62.29%	68.57%
Positive Predictive Value (PPV)	61.75%	67.21%
Negative Predictive Value (NPV)	62.87%	68.86%
F1 Score	0.63.14	0.6869
Positive Likelihood Ratio (LR+)	1.614	2.050
Negative Likelihood Ratio (LR−)	0.5905	0.4513

Diagnostic metrics including sensitivity, specificity, accuracy, predictive values, F1 score, and likelihood ratios are compared between ORAD III and ORAD DDx.

DISCUSSION

In this study, ORADIII and ORAD DDx systems were evaluated against the gold standard diagnosis i.e., histopathology diagnosis.

Concordance between the histopathological diagnosis and ORADIII was 90 (51.42%), partial concordance in 23 (13.14%), and discordance in 62 (35%). Whereas, concordance between the histopathological diagnosis and ORAD DDx was 116 (66.29%), partial concordance in 7 (4%), and discordance in 52 (29.71%) among the study sample.

A Previous study reported concordance between histopathology diagnoses with ORAD was 46% and CHAT-GPT was 41%. However, partial concordance was higher, and discordance was lower for ORAD (35% and 19%) and CHAT-GPT (50% and 8%), which was opposite to the findings in comparison to the present study.[44]

The present study reported the sensitivity and specificity of ORADIII against the gold standard to be 64.57% and 60%, respectively. A 15-year retrospective New Zealand-based study comparing the clinicopathological concordance of clinicians, Chat-GPT4, and ORAD for odontogenic keratocysts and odontogenic tumors reported sensitivity of ORAD at 66.67% and specificity at 84.94 %. [44]

Sensitivity is a measure of true positive (TP) tests in comparison to the gold standard test.[37] The present study showed that the ability of ORADIII to yield a positive outcome for a subject who has that disease/lesion/pathology is less than reported by the previous study.[44]

Specificity is the measure of true negative (TN) tests in comparison to the gold standard test, which does not have a disease/lesion/pathology.[37] The present study observed that ORADIII's ability to report the absence of a lesion in normal data was low in comparison to previous reported studies.[44]

This could be attributed to the sample selection. Since Kim et. al used a sample consisting of only orthokeratinised cysts and odontogenic tumors, whereas in the present study, all intra-bony jaw lesions were considered for Lesion Group and normal anatomic landmarks and variations for the Non Lesion Group .[44]

The diagnostic accuracy of ORADIII was 62.29%. This result is higher than a 2024 study conducted by Kalambe et. al, comparing diagnostic accuracy of ORAD with histopathological diagnosis. They reported diagnostic accuracy of ORAD as 50%, whereas that of maxillofacial radiologists was 68.4%.[36] Another preliminary study done in 2011, by Simoes et. al., reported that 67% of the ORADs' generated results did not match the biopsy reports.[47]

The former study included only 38 OPGs, whereas the latter was done using sample of nine confirmed histopathology diagnoses only.[36, 47]

Whereas the Kim et al study reported the diagnostic accuracy of ORAD as 81.68%, which is quite high, could be due to the fact that only OKCS and odontogenic tumors were included.[44]

ORAD DDx, which is relatively new, showed slightly higher sensitivity, specificity, and diagnostic accuracy than ORADIII.

This can be attributed to the fact that ORAD DDx gives its diagnoses under various categories such as anatomic, inflammatory, cyst, tumor and so on, so researcher can

choose top four from any category. This is not so with ORADIII, which gives its diagnosis, considering the percentage of probability.

The F1 score for ORADIII in the present study was 0.6314. This is slightly higher than a previous study (Kim et al), which reported 56.47% (0.5647) for ORAD. [44]

When testing machine learning/AI, F1 score is preferred over accuracy since accuracy is the percentage of right answers, and it gives a general sense of the quality of the model, potentially overlooking many distinctions of the situation. Precision is the percentage of non-false positives: a low value means that the model is giving many false positives. Recall is the percentage of positives missed by the model. Precision and recall are competing metrics, difficult to adjust concurrently. To increase the latter, the model should be made more sensitive. This change, however, would cause the model to also generate more false alarms, which would, by definition, reduce the precision and vice versa. This motivates the introduction of the metric, F1 Score. It is a combination (harmonic mean) of precision and recall, which makes it an informative summary of the quality of the model. The F1 score ranges between 0 and 1, with 0 denoting the lowest possible result and 1 denoting a flawless result, meaning that the model accurately predicted each label.[48] The reported range falls in the OK/average category.[49]

The Positive Predictive Value (PPV) and Negative Predictive Value (NPV) for ORADIII were 61.75% and 62.87%, respectively, as compared to 50.62% and 94.14% reported in a previous study.[44]

PPVs determine that out of all of the positive findings reported by the test, how many are true positives; similarly, NPVs determine, out of all of the negative findings, how many are true negatives. The value increases toward 100 as it approaches a ‘gold

standard'.[37*] Interpreting our result in comparison to a previous study, though PPV is in a similar range, the NPV of the previous study was high, near perfect, which could again be attributed to the choice of lesion/pathology tested for.[44]

For ORAD DDX in the present study, the F1 score fell in the same range, but PPV and NPV were somewhat higher than ORADIII.

The Positive Likelihood Ratio (LR+) and Negative Likelihood Ratio (LR-) for ORADIII were 1.614 and 0.5905, respectively. Whereas, the Positive Likelihood Ratio (LR+) and Negative Likelihood Ratio (LR-) for ORAD DDx were 2.050 and 0.4513, respectively, in present study sample.

The LR is the probability of a given test result in a patient result in a person without the target disorder. The components of the LR are calculated vertically, and like the sensitivity and specificity, are immune to the prevalence. The LR of a positive test result: LR+ is the probability that an individual with the target disorder has a positive test probability than an individual without the target disorder has a positive test.[50]

When interpreting the values from the present test, it falls in the category of minimal increase in the likelihood of disease.[50] So represents low likelihood ratios, which are similar for both the CDSSs.

LR- is the probability that an individual with the condition has a negative test/probability in comparison to an individual without the condition has a negative test.[50]

The reduced rate for the present study could be attributed to the inclusion of the majority of intrabony lesions, since they have variable radiographic presentation, be it

multilocular or unilocular, or overlapping of features of different entities, since OPG is a two-dimensional study.

Also, prevalence consideration in ORADIII and common in ORAD DDx was marked when searching the differential. This could also generate region-specific diagnoses and which may not be as prevalent in our selected sample.

CONCLUSION

In the present study, both ORADIII and ORAD DDx were evaluated against histopathological diagnosis, which served as the gold standard. The diagnostic performance of ORADIII and ORAD DDx was found to be moderate, with ORAD DDx performing slightly better in most diagnostic metrics.

The sensitivity and specificity of ORADIII in our study were 64.57% and 60%, respectively, while ORAD DDx showed 70.28% sensitivity and 65.71% specificity, suggesting modest performance in identifying true positives and true negatives. The diagnostic accuracy of ORADIII (62.29%) and ORAD DDx (68.57%) also indicates room for improvement, especially when compared to histopathological confirmation.

Regarding the level of diagnosis, both ORADIII and ORAD DDx provide radiographic differential diagnoses based on pattern recognition and feature input. However, they do not provide histological subtyping or cellular-level characterization, which is only possible through histopathology. Thus, the CDSSs operate at a radiographic diagnostic level, not at the definitive or tissue-based diagnostic level.

In summary, the current study demonstrates that ORAD DDx outperforms ORADIII modestly in diagnostic agreement with histopathological diagnosis, particularly in terms of sensitivity, specificity, and concordance. However, both tools should be used as adjuncts to clinical and radiological evaluation, not as standalone diagnostic systems.

LIMITATIONS

- In the present study a heterogenous group of large variety of intra bony large lesions were included, this may have reduced the precision of both CDSS.
- This study had small sample size and due to retrospective nature of the study, clinical features could not be included.
- As the samples of the study were collected from patients visiting Department of Oral Medicine and Radiology, Dhulikhel Hospital, the result cannot be generalized to the larger population group.
- Since convenience sampling method was used, there can in selection bias in particularly for non lesion group.
- The study was based on two-dimensional OPG with manual input. So, it is prone to errors.
- The inter-observer reliability/variability test was not done.

RECOMMENDATIONS

- Separate evaluations of the CDSS for specific lesion types (e.g., odontogenic tumors, cysts, fibro-osseous lesions) would help in assessing performance within more homogenous groups and reduce interpretive variability.
- Future studies should consider using Cone Beam Computed Tomography (CBCT) or other 3D imaging modalities alongside OPGs to capture more detailed lesion characteristics, allowing for better evaluation by both clinicians and CDSS.
- Prospective studies would be able to include clinical data such as pain, swellings and so on.
- Developers of ORADIII and ORAD DDx should consider building region-specific versions or modules that reflect the prevalence and radiologic appearance of pathologies commonly encountered in different populations.
- Further validation should be conducted using larger sample sizes from multiple geographic locations to improve generalizability and assess system performance across diverse populations and settings.

REFERENCES

1. Ding H, Wu J, Zhao W, Matinlinna JP, Burrow MF, Tsoi JK. Artificial intelligence in dentistry—A review. Front Dent Med. 2023;4(1):10851-52.

2. Hung K, Montalvao C, Tanaka R, Kawai T, Bornstein MM. The use and performance of artificial intelligence applications in dental and maxillofacial radiology: A systematic review. Dentomaxillofac Radiol. 2020;49(1):201901-07.

3. Hung KF, Yeung AWK, Bornstein MM, Schwendicke F. Personalized dental medicine, artificial intelligence, and their relevance for dentomaxillofacial imaging. Dentomaxillofac Radiol. 2023;52(1):202203-35.

4. Heo MS, Kim JE, Hwang JJ, Han SS, Kim JS, Yi WJ et al. Artificial intelligence in oral and maxillofacial radiology: what is currently possible?. Dentomaxillofac Radiol. 2021;50(3):202003-75.

5. Calazans MA, Ferreira FA, Alcoforado MD, Santos AD, Pontual AD, Madeiro F. Automatic classification system for periapical lesions in cone-beam computed tomography. Sens. 2022;22(17):6481.

6. Hirosawa T, Harada Y, Yokose M, Sakamoto T, Kawamura R, Shimizu T. Diagnostic accuracy of differential-diagnosis lists generated by generative pretrained transformer 3 chatbot for clinical vignettes with common chief complaints: a pilot study. Int J Environ Res Public Health. 2023;20(4):3378.

7. Shankar, P Ravi. (2022). Artificial intelligence in healthcare in Nepal. Janaki Medical College Journal of Medical Science. 10. 10.3126/jmcjms.v10i1.44622.

8. Upadhayay, A. R. (2024). Artificial Intelligence: Opportunities and Challenges for the Global South with a Focus on Nepal. *NCWA Annual Journal*, *55*(01), 44–52.

9. Wellwood JM, Spiegelhalter DJ. Computers and the diagnosis of acute abdominal pain. Br J Hosp Med. 1989;41(6):564–7.

10. Hatton GE, Pedroza C, Kao LS. Bayesian Statistics for Surgical Decision Making. Surg Infect. 2021;22(6):620–5.

11. Wiener F, Laufer D, Ribak A. Computer-aided diagnosis of odontogenic lesions. Int J Oral Maxillofac Surg. 1986;15(5):592–6.

12. White SC. Computer-aided differential diagnosis of oral radiographic lesions. Dentomaxillofac Radiol. 1989;18(2):53–9.

13. Iwasaki H. Bayesian belief network analysis applied to determine the progression of temporomandibular disorders using MRI. Dentomaxillofac Radiol. 2015;44(4):20140279.

14. Hamilton B. DXplain Mobile: An Assessment of a Smartphone-Based Expert Diagnostic System. Harvard University; 2016.

15. Castaneda C, Nalley K, Mannion C, Bhattacharyya P, Blake P, Pecora A, et al. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. J Clin Bioinforma. 2015;5(1):1-6.

16. Dinevski D, Bele U, Šarenac T, Rajkovič U, Šušteršič O. Clinical decision support systems. Telemedicine techniques and applications. 2011:185-210.

17. Kho CLW, Chow DY, Wong JM, Loh JW, Sim YF, Gan MJS, et al. Comparing the effectiveness of two diagnostic approaches for the interpretation of oral radiographic lesions by dental students. Adv Health Sci Educ. 2023;28(1):107–26.

18. Goh WP, Tao X, Zhang J, Yong J. Decision support systems for adoption in dental clinics: a survey. Knowledge-Based Systems. 2016 Jul 15;104(1):195-206.

19. White SC, Pharoah MJ. White and Pharoah's oral radiology: principles and interpretation. Mosby; 2008.

20. Scarfe WC, Toghyani S, Azevedo B. Imaging of benign odontogenic lesions. Radiologic Clinics. 2018;56(1):45-62.

21. Soluk-Tekkesin M, Wright JM. The World Health Organisation classification of odontogenic lesions: a summary of the changes of the 2022 (5^{th}) edition. Turk J Pathol. 2022; 38(2): 168.

22. Wood NK, Goaz PW. Differential Diagnosis of Oral and Maxillofacial Lesions. Mosby; 1997.

23. Bajracharya D, Ojha B, Koju S, Maharjan N. Clinico pathological study of Odontogenic cysts and tumors in a Tertiary care Dental hospital of Nepal. J Kathmandu Med Coll. 2022;11(1):52–7.

24. Mosier KM. Lesions of the Jaw. Semin Ultrasound CT MRI. 2015;36(5):444–50.

25. Almazrooa S, Binmadi NO, Khalifa HM, Jadu FM, Jan AM, Meisha DE. The Agreement Rate between Radiographic Interpretation and Histopathologic Diagnosis of Jaw Lesions. Radiol Res Pract. 2019;2019:1–5.

26. Bajracharya B, Poudel P, Bajracharya D, Bhattacharyya S, Shakya P. Eosinophilic Granuloma of Mandible: A Diagnostic Challenge. Kathmandu Univ Med J. 2018;62(2):201-3.

27. Miloglu O, Torenek-Agirman K, Dalci HL, Miloglu FD, Yildizbas Z. Diagnosis of common intraosseous lesions of the dentomaxillofacial region by chemometry-assisted FT-IR spectroscopy in dental tissue samples. J Stomatol Oral Maxillofac Surg. 2024;125(3):101706.

28. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017;2(4):230-43.

29. Chen Z, Liang N, Zhang H, Li H, Yang Y, Zong X, et al. Harnessing the power of clinical decision support systems: challenges and opportunities. Open Heart. 2023;10(2):2432.

30. Mendonça EA. Clinical Decision Support Systems: Perspectives in Dentistry. J Dent Educ. 2004;68(6):589–97.

31. Vikram K, Karjodkar FR. Decision Support Systems in Dental Decision Making: An Introduction. J Evid Based Dent Pract. 2009;9(2):73–6.

32. Tandon D, Rajawat J, Banerjee M. Present and future of artificial intelligence in dentistry. J Oral Biol Craniofacial Res. 2020;10(4):391–6.

33. Van Baalen S, Boon M, Verhoef P. From clinical decision support to clinical reasoning support systems. J Eval Clin Pract. 2021;27(3):520–8.

34. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. Npj Digit Med. 2020;3(1):17-22.

35. Shuja AT, Nellengara AA, R SS, Mani HP, Joy J. Artificial intelligence in oral medical radiology & clinical decision support systems (CDSS). Int J Nov Res Dev. 2023;8(3):551–7.

36. Kalambe YM, Bhadage CJ, Bhoosreddy AR, Virkar VV, Santan RS, Shahakar MS. Comparison of diagnostic accuracy of ORAD an artificial intelligence powered software and maxillofacial radiologist in the diagnosis of benign lesions of the jaw–a cross-sectional study. Int J Dent Med Sci Res. 2024;6(3):348–52.

37. Shreffler J, Huecker MR. Diagnostic testing accuracy: Sensitivity, specificity, predictive values and likelihood ratios. 2020.

38. Šimundić AM. Measures of Diagnostic Accuracy: Basic Definitions. Med Biol Sci. 2009;19(4):203–11.

39. McGenity C, Clarke EL, Jennings C, Matthews G, Cartlidge C, Freduah-Agyemang H, Stocken DD, Treanor D. Artificial intelligence in digital pathology: a diagnostic test accuracy systematic review and meta-analysis. arXiv. 2023;1(1):1-26.

40. Nafees A, Khan M, Chow R, Fazelzad R, Hope A, Liu G, et al. Evaluation of clinical decision support systems in oncology: An updated systematic review. Crit Rev Oncol Hematol. 2023;192(1):1041-43.

41. Mazo C, Kearns C, Mooney C, Gallagher WM. Clinical Decision Support Systems in Breast Cancer: A Systematic Review. Cancers. 2020;12(2):369.

42. Khalfallah HB, Jelassi M, Demongeot J, Saoud NBB. Decision support systems in healthcare: systematic review, meta-analysis and prediction, with example of COVID-19. AIMS Bioeng. 2023;10(1):27–52.

43. Mago VK, Bhatia N, Bhatia A, Mago A. Clinical decision support system for dental treatment. J Comput Sci. 2012;3(5):254–61.

44. Kim P, Seo B, De Silva H. Concordance of clinician, Chat-GPT4, and ORAD diagnoses against histopathology in Odontogenic Keratocysts and tumours: a 15-Year New Zealand retrospective study. Oral Maxillofac Surg. 2024;28(4):1557–69.

45. Cosson J. Interpreting an orthopantomogram. Aust J Gen Pract. 2020;49(9):550–5.

46. Soluk-tekkesin M, Wright JM. The world health organization classification of odontogenic lesions: a summary of the changes of the 2017 (4th) edition. Turk J Pathol. 2018;34(1):1-8.

47. Simões AF, Correia A, Marques T, Figueiredo R. Preliminary study of the clinical application of the Clinical Decision Support System ORAD II in a university dental clinic. Computational Vision and Medical Image Processing: VipIMAGE 2011. 2011;12(2):93-9.

48. Conciatori M, Valletta A, Segalini A. Improving the quality evaluation process of machine learning algorithms applied to landslide time series analysis. Comput Geosci. 2024:184(1):105531.

49. Medhat S, Abdel-Galil H, Aboutabl AE, Saleh H. Iterative magnitude pruning-based light-version of AlexNet for skin cancer classification. Neural Comput and Appl. 2024;36(3):1413-28.

50. Parikh R, Parikh S, Arun E, Thomas R. Likelihood ratios: Clinical application in day-to day practice. Indian J Ophthalmol. 2009;57(3):217-23.

ANNEXURES

Kathmandu University School of Medical Sciences

TOPIC: Diagnostic Accuracy of Clinical Decision Support Systems (CDSSs) ORADIII and ORAD DDx to histopathological diagnosis of jaw lesions.

CONDUCTED BY: Dr. Harleen Bali

1. Participant Information

- Participant ID/Study No:
- Age:
- Gender:

---

2. RADIOGRAPHIC FEATURES

a. Location

1. Which jaw contains the lesion?
2. The lesion center is in what region ?
3. The relationship of the lesion to teeth is:
4. Please estimate the number of lesions:
5. What is the maximum size of the lesion?
6. Where is the origin of the lesion?

b. Periphery

1. The borders of the lesion are:
2. The loculation of the lesion is:

c. Internal Structure

1. The contents of the lesions are:
2. Does the lesion contain one or more teeth ?

d. Effects on Surrounding Structures
    1. Does the lesion expand the bony cortex?
    2. Does the lesion cause root resorption ?
    3. Does the lesion cause tooth displacement or impaction?

Date:
Signature of Principal Investigator:

KATHMANDU UNIVERSITY
SCHOOL OF MEDICAL SCIENCES

November 06, 2024

To,

Dr. Harleen Bali
Kathmandu University School of Medical Sciences
Dhulikhel, Kavre.

Subject: Approval of Research Proposal

Dear Dr. Harleen Bali

This is to certify that the following thesis (Masters in Medical Research) protocol and related documents have been reviewed and granted approval by Institutional Review Committee, Kathmandu University School of Medical Sciences (IRC, KUSMS) for implementation on 29 October 2024.

IRC-KUSMS Approval No.	**303/24**	Duration of Approval	November, 2025
Principal Investigator (s)	Dr. Harleen Bali	Sponsor Institute	N/A
Title	**"Diagnostic Accuracy of Clinical Decision Support Systems (CDSSs) ORADIII and ORAD DDx to histopathological diagnosis of jaw lesions"**		
Other Members of Research Team (Co- Investigators)	Dr. Dashrath Kafle, Dr. Sagar Adhikari, Dr. Nitesh Kumar Chaurasia, Dr. Pratibha Poudel, Dr. Bhoj Raj Adhikari.		
IRC-KUSMS, Administrative fee	NRs. 500.00		
Chairperson of IRC-KUSMS	**Name**		
	Prof. Dr. Prabodh Risal		
Investigator Responsibilities: ➢ Comply with all relevant International and NHRC ethical guidelines. ➢ Submit final report after completion of protocol at IRC-KUSMS.			

If you have any questions, please contact the IRC-KUSMS section at Kathmandu University School of Medical Sciences/ Kathmandu University Hospital.

With best regards,

Dr. Dipesh Tamrakar
Member Secretary, IRC-KUSMS

Dhulikhel, Kavre GPO Box 11008 Kathmandu, Nepal Phone: 00977 11 490497 Fax: 00977 11 490707 Email: kusms@ku.edu.np www.dhulikhelhospital.org Collaborative Program of Kathmandu University and Dhulikhel Hospital