

Supplementary File 1

Questionnaire

Comparative Evaluation of Synthetic and Real-World Data in Predicting Oral Premalignant Lesions: A Machine Learning Approach from Rural India

Submission ID: 0a85331e-1f8f-4fff-b142-fd7e3c6aa534

Instrument Title	Structured Oral Health Screening and Risk Assessment Questionnaire for Oral Premalignant Lesion Detection
Setting	Community-based, house-to-house survey across 100 villages, Hassan District, Karnataka, India
Development year	2019
Previously published	No — developed specifically for this study
Mode of administration	Structured personal interview by trained dental clinician + direct clinical oral examination
Language	English (field administration in Kannada; this file presents the English-language version)
Total items	72 items across 6 sections (Sections A–F)
Ethics approval	Institute Ethics Committee, Hassan Institute of Medical Sciences (IEC/HIMS/RR72/21-05-2019)
Informed consent	Written informed consent obtained from all participants prior to administration

Instructions to interviewer: Read each question aloud clearly in the participant's preferred language. Circle or record the participant's response. Do not prompt or suggest responses. For clinical examination items (Section F), record findings directly from oral inspection. Estimated administration time: 20–25 minutes.

Instructions to participant: This questionnaire is part of a health survey to understand oral health in your community. All information is strictly confidential. There are no right or wrong answers — please respond honestly. Participation is voluntary and you may decline any question.

SECTION A: Sociodemographic Information (Items 1–12)

To be completed by the interviewer based on participant's responses.

Q No.	Question / Item	Response Type	Response Options / Format
1.	What is your age (in completed years)?	Open numeric	Record in years: ____
2.	What is your gender?	Single choice	1 = Male 2 = Female 3 = Other
3.	What is your religion?	Single choice	1 = Hindu 2 = Muslim 3 = Christian 4 = Other (specify): ____
4.	What is your highest level of education completed?	Single choice	1 = No formal education 2 = Primary 3 = Secondary 4 = Pre-University 5 = Graduate & above

Q No.	Question / Item	Response Type	Response Options / Format
5.	What is your current marital status?	Single choice	1 = Unmarried 2 = Married 3 = Widowed 4 = Separated/Divorced
6.	What is your primary occupation?	Single choice	1 = Agriculture 2 = Manual labour 3 = Skilled/Trade 4 = Business 5 = Homemaker 6 = Student 7 = Government/Professional 8 = Unemployed 9 = Other: ___
7.	What type of family do you live in?	Single choice	1 = Nuclear 2 = Joint 3 = Extended
8.	How many people live in your household (including yourself)?	Open numeric	Count: ___
9.	What is your household's total monthly income (INR)?	Open numeric	Amount in INR: ___
10.	What is your socioeconomic status? (Interviewer: use Modified Kuppaswamy Scale)	Single choice	1 = Upper 2 = Upper-middle 3 = Middle 4 = Lower-middle 5 = Lower
11.	What is the name of your village/town?	Open text	Record: ___
12.	What is the distance from your home to the nearest Primary Health Centre (km)?	Open numeric	Distance in km: ___

SECTION B: Family History and Household Habits (Items 13–24)

Ask about the participant's immediate family members (parents, siblings, spouse, children) living in the same household or with shared habits.

Q No.	Question / Item	Response Type	Response Options / Format
13.	Has any blood relative (parent, sibling, child) been diagnosed with oral cancer?	Single choice	0 = No 1 = Yes 9 = Don't know
14.	If yes (Q13), who was affected?	Multiple choice	1 = Parent 2 = Sibling 3 = Child 4 = Other: ___ (Skip if Q13 = No)
15.	Do any members of your household use tobacco in any form?	Single choice	0 = No 1 = Yes
16.	If yes (Q15), what type of tobacco do family members use?	Multiple choice	1 = Cigarette 2 = Bidi 3 = Chewing tobacco 4 = Gutka 5 = Betel quid 6 = Other: ___ (Skip if Q15 = No)
17.	Have any family members complained of any problem in their oral cavity in the past 12 months?	Single choice	0 = No 1 = Yes
18.	If yes (Q17), what type of complaint?	Multiple choice	1 = Mouth sore 2 = White patch 3 = Difficulty opening mouth 4 = Burning sensation 5 = Bleeding

Q No.	Question / Item	Response Type	Response Options / Format
			gums 6 = Other: ____ (Skip if Q17 = No)
19.	Has any family member sought treatment for an oral problem in the past 12 months?	Single choice	0 = No 1 = Yes 9 = Don't know
20.	Do any family members consume alcohol regularly?	Single choice	0 = No 1 = Yes
21.	Has any family member been diagnosed with any type of cancer (other than oral)?	Single choice	0 = No 1 = Yes 9 = Don't know
22.	If yes (Q21), what type of cancer?	Open text	Record: ____ (Skip if Q21 = No)
23.	Do you have a family history of diabetes or hypertension?	Multiple choice	0 = Neither 1 = Diabetes 2 = Hypertension 3 = Both
24.	Are you currently taking any long-term medications?	Single choice	0 = No 1 = Yes — specify: ____

SECTION C: Tobacco, Chewing, and Alcohol Use Habits (Items 25–42)

Record participant's personal habits. For each habit, capture current status, type, frequency, duration, and quit history.

Q No.	Question / Item	Response Type	Response Options / Format
25.	Have you ever smoked cigarettes, bidis, or any other tobacco product?	Single choice	0 = Never 1 = Current smoker 2 = Ex-smoker (quit)
26.	If ever smoked (Q25), what type do/did you primarily smoke?	Single choice	1 = Cigarette 2 = Bidi 3 = Hookah/Pipe 4 = Cigar 5 = Other: ____ (Skip if Q25 = Never)
27.	How many cigarettes/bidis do/did you smoke per day (on average)?	Open numeric	Count per day: ____ (Skip if Q25 = Never)
28.	For how many years have/did you smoke?	Open numeric	Years: ____ (Skip if Q25 = Never)
29.	If ex-smoker (Q25 = 2), how many years ago did you quit?	Open numeric	Years since quitting: ____ (Skip if not ex-smoker)
30.	Do you currently chew any tobacco product (gutka, pan masala, khaini, tobacco with betel leaf)?	Single choice	0 = Never 1 = Current user 2 = Ex-user (quit)
31.	If ever chewed (Q30), what type do/did you chew?	Multiple choice	1 = Gutka 2 = Pan masala 3 = Khaini 4 = Tobacco with betel leaf 5 = Areca nut alone 6 = Other: ____ (Skip if Q30 = Never)
32.	How many times per day do/did you chew (on average)?	Open numeric	Times per day: ____ (Skip if Q30 = Never)
33.	For how many years have/did you chew?	Open numeric	Years: ____ (Skip if Q30 = Never)

Q No.	Question / Item	Response Type	Response Options / Format
34.	Do you keep the chewed quid/tobacco in your mouth for an extended period without spitting?	Single choice	0 = No 1 = Yes — usual duration kept: ___ minutes (Skip if Q30 = Never)
35.	At what site do you usually keep the quid in the mouth?	Single choice	1 = Buccal mucosa (cheek) 2 = Under tongue 3 = Vestibule 4 = Varies (Skip if Q34 = No)
36.	Do you consume betel quid (paan) without tobacco?	Single choice	0 = No 1 = Yes — frequency: ___/day
37.	Do you consume areca nut (supari) in any form?	Single choice	0 = No 1 = Yes — frequency: ___/day
38.	Do you consume alcohol?	Single choice	0 = Never 1 = Current user 2 = Ex-user (quit)
39.	If yes (Q38), what type of alcohol do you primarily consume?	Single choice	1 = Toddy 2 = Arrack/local brew 3 = Beer 4 = Whisky/spirits 5 = Other: ___ (Skip if Q38 = Never)
40.	How frequently do you consume alcohol?	Single choice	1 = Daily 2 = Weekly 3 = Occasionally (< once/week) (Skip if Q38 = Never)
41.	How many years have you been consuming alcohol?	Open numeric	Years: ___ (Skip if Q38 = Never)
42.	Do you combine chewing and smoking at the same time?	Single choice	0 = No 1 = Yes 2 = Not applicable

SECTION D: Oral Hygiene Knowledge and Practices (Items 43–58)

Assess the participant's knowledge, attitudes, and self-care behaviours related to oral health.

Q No.	Question / Item	Response Type	Response Options / Format
43.	Are you aware that tobacco use can cause oral cancer?	Single choice	0 = No 1 = Yes, somewhat 2 = Yes, fully aware
44.	Are you aware of what oral cancer looks like (e.g., white or red patches, non-healing ulcers)?	Single choice	0 = No 1 = Yes, partially 2 = Yes, clearly
45.	Do you know how to perform a self-examination of your mouth for early warning signs?	Single choice	0 = No 1 = Yes — how often do you do it? ___/month
46.	Have you ever performed a self-oral visual inspection in the past 6 months?	Single choice	0 = No 1 = Yes — last time: ___
47.	How many times do you brush your teeth per day?	Single choice	0 = Never 1 = Once daily 2 = Twice daily 3 = After every meal
48.	What do you use to clean your teeth?	Multiple choice	1 = Toothbrush + paste 2 = Finger + powder 3 = Neem stick 4 = Charcoal 5 = Nothing 6 = Other: ___

Q No.	Question / Item	Response Type	Response Options / Format
49.	Do you rinse your mouth with mouthwash regularly?	Single choice	0 = No 1 = Yes — frequency: ___/week
50.	Do you use any traditional tooth-cleaning substance (e.g., ash, salt)?	Single choice	0 = No 1 = Yes — specify: ___
51.	Do you consume excessively spicy food on a daily basis?	Single choice	0 = No 1 = Occasionally 2 = Yes, regularly
52.	Do you have any sharp, broken, or irregular teeth that scratch your cheek or tongue?	Single choice	0 = No 1 = Yes — for how long? ___ months
53.	Do you wear dentures? If yes, do they fit properly?	Single choice	0 = No dentures 1 = Dentures, well-fitting 2 = Dentures, ill-fitting
54.	Have you ever had any dental or oral treatment in the past?	Single choice	0 = No 1 = Yes — type of treatment: ___
55.	How long ago was your last visit to a dentist or oral health professional?	Single choice	1 = Within 6 months 2 = 6–12 months 3 = 1–3 years 4 = More than 3 years 5 = Never visited
56.	What was the reason for your last dental visit?	Single choice	1 = Routine check-up 2 = Pain/complaint 3 = Extraction 4 = Treatment for lesion 5 = Never visited (Skip if Q55 = Never)
57.	Have you ever been told by a doctor or dentist that you have a lesion or problem in your mouth?	Single choice	0 = No 1 = Yes — specify: ___
58.	Are you aware of any community oral health program or screening camp in your area?	Single choice	0 = No 1 = Yes — have you attended? 0 = No 1 = Yes

SECTION E: Symptoms and Self-Reported Oral Complaints (Items 59–68)

Ask the participant about any oral symptoms experienced in the past 3 months. Do not prompt with diagnoses.

Q No.	Question / Item	Response Type	Response Options / Format
59.	Have you noticed any sore, ulcer, or wound in your mouth that has not healed within 2 weeks?	Single choice	0 = No 1 = Yes — duration: ___ weeks location: ___
60.	Have you noticed any white patch, thickening, or roughness inside your mouth?	Single choice	0 = No 1 = Yes — location: ___
61.	Have you noticed any red patch or red and white mixed patch in your mouth?	Single choice	0 = No 1 = Yes — location: ___
62.	Do you experience a burning sensation in your mouth?	Single choice	0 = No 1 = Occasionally 2 = Frequently
63.	Do you have difficulty opening your mouth fully (reduced mouth opening)?	Single choice	0 = No 1 = Mild — can open with some effort 2 = Severe — very limited opening

Q No.	Question / Item	Response Type	Response Options / Format
64.	If yes (Q63), for how long have you had difficulty opening your mouth?	Open numeric	Duration in months: ____ (Skip if Q63 = No)
65.	Do your gums bleed when you brush, eat, or spontaneously?	Single choice	0 = No 1 = Sometimes 2 = Frequently / recurrently
66.	Do you have difficulty swallowing food or liquids?	Single choice	0 = No 1 = Occasionally 2 = Frequently
67.	Do you experience persistent numbness or tingling in your mouth, lips, or tongue?	Single choice	0 = No 1 = Yes — location: ____
68.	Have you experienced unexplained weight loss (more than 5 kg) in the past 6 months?	Single choice	0 = No 1 = Yes

SECTION F: Clinical Oral Examination Findings (Items 69–72)

To be completed by the examining Oral Medicine specialist following direct clinical oral examination under adequate lighting. Do not rely on participant self-report for this section.

Q No.	Question / Item	Response Type	Response Options / Format
69.	Is there any clinically visible lesion identified in the oral cavity on examination?	Single choice	0 = No lesion detected 1 = Lesion present — specify type and site: ____
70.	Is there clinical evidence of difficulty in opening the mouth (trismus) on examination? (Assess inter-incisal distance)	Single choice	0 = Normal (≥ 35 mm) 1 = Mild restriction (20–34 mm) 2 = Severe restriction (< 20 mm) Record inter-incisal distance: ____ mm
71.	Is there evidence of recurrent gingival bleeding on probing or observed spontaneously during examination?	Single choice	0 = No bleeding 1 = Bleeding on probing 2 = Spontaneous / recurrent bleeding
72.	Is there any red or white mucosal patch identified on clinical examination? (Classify as leukoplakia, erythroplakia, or OSMF based on WHO criteria)	Single choice	0 = No mucosal patch 1 = White patch (leukoplakia suspect) 2 = Red patch (erythroplakia suspect) 3 = Mixed red-white patch 4 = OSMF features Clinician's provisional diagnosis: ____

OUTCOME CLASSIFICATION (Clinician to complete)

OPL Status (Primary Outcome):

OPL-Negative OPL-Positive
(Based on clinical consensus of Sections E & F)

Provisional Diagnosis (if OPL-Positive):

Leukoplakia OSMF Erythroplakia
Other: ____
Referral recommended: Yes No

Examining Clinician: _____ Signature: _____ Date: ____/____/____

Note on ML feature extraction: Of the 72 items, 24 predictor variables were extracted for machine learning analysis: age (Q1), gender (Q2), religion (Q3), occupation (Q6), family type (Q7), family size (Q8), total monthly income (Q9), socioeconomic status (Q10), family history of oral cancer (Q13), habits among family members (Q15), family complaints regarding the oral cavity (Q17), family oral cavity complaints (Q18/Q19), smoking status (Q25), habit of chewing (Q30), habit of keeping quid in mouth (Q34), alcohol consumption (Q38), knowledge about oral hygiene (Q43–Q44), mouth-washing practices (Q49), excessive spicy food intake (Q51), sharp teeth or ill-fitted dentures (Q52/Q53), any prior oral treatment (Q54), performance of self-oral visual inspection (Q46), presence of any visible lesion (Q69), difficulty in opening the mouth/trismus (Q70), recurrent gingival bleeding (Q71), and red or white mucosal patch (Q72). The primary binary outcome variable (OPL positive/negative) was determined by specialist clinical consensus based on Sections E and F findings.