PROFESSOR ALEXANDER ROSS KERR (Orcid ID : 0000-0002-2162-8572)
DR SOK CHING CHEONG (Orcid ID : 0000-0002-6765-9542)

Article type : Original Article

Image collection and annotation platforms to establish a multi-source database of oral lesions

Senthilmani Rajendran[a], Jian Han Lim[b], Kohgulakuhan Yogalingam[a], Thomas George Kallarakkal[c,d], Rosnah Binti Zain[c,d,e], Ruwan Duminda Jayasinghe[f], Jyotsna Rimal[g], Alexander Ross Kerr[h], Rahmi Amtha[i], Karthikeya Patil[j], Roshan Alex Welikala[k], Ying Zhi Lim[a], Paolo Remagnino[k], John Gibson[p], Wanninayake Mudiyanselage Tilakaratne[c,f], Chee Sun Liew[l,m,n], Yi-Hsin Yang[o], Sarah Ann Barman[k], Chee Seng Chan[b] and Sok Ching Cheong[a,c]

[a]Digital Health Research Unit, Cancer Research Malaysia, 47500 Subang Jaya, Malaysia.
[b]Centre of Image and Signal Processing, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia.
[c]Department of Oral and Maxillofacial Clinical Sciences, Faculty of Dentistry, University of Malaya, 50603 Kuala Lumpur, Malaysia.
[d]Oral Cancer Research and Coordinating Centre (OCRCC), Faculty of Dentistry, University of Malaya, 50603 Kuala Lumpur, Malaysia.
[e]Faculty of Dentistry, MAHSA University, Bandar Saujana Putra, 42610 Jenjarom, Malaysia.
[f]Centre for Research in Oral Cancer, Department of Oral Medicine and Periodontology, Faculty of Dental Sciences, University of Peradeniya, Peradeniya, 20400, Sri Lanka.
[g]Department of Oral Medicine and Radiology, BP Koirala Institute of Health Sciences, Dharan, 56700, Nepal.
[h]Oral and Maxillofacial Pathology, Radiology and Medicine, New York University, New York, NY 10010, USA.
[i]Faculty of Dentistry, Trisakti University, Kota Jakarta Barat, Jakarta 11440, Indonesia.

[j]Oral Medicine and Radiology, JSS Dental College and Hospital, JSS Academy of Higher Education & Research, Mysuru 570 015 Karnataka, India.
[k]Digital Information Research Centre, Faculty of Science, Engineering and Computing, Kingston University, Surrey, KT1 2EE, United Kingdom.
[l]Department of Computer System & Technology, Faculty of Computer Science & Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia
[m]Centre of Data Analytics, Research Management & Innovation Complex, University of Malaya, 50603 Kuala Lumpur, Malaysia
[n]Centre for Data Science, University of Malaya, 50603 Kuala Lumpur, Malaysia
[o]National Institute of Cancer Research, National Health Research Institutes, Tainan 70456, Taiwan
[p]Institute of Dentistry, University of Aberdeen, Aberdeen, AB25 2ZR, United Kingdom

**** Corresponding author**
Sok Ching Cheong
Digital Health Research Unit, Cancer Research Malaysia, 47500 Subang Jaya, Selangor, Malaysia.
Tel: +603 5650 9797 Fax: +603 5650 9787
Email address: sokching.cheong@cancerresearch.my

Running Head: Oral lesion database

Key Words: Oral lesion image database, Annotation Tool, Oral Cancer, Oral Potentially Malignant Disorders, Access to Care

Preprint
A preprint of this article has been uploaded to Research Square (DOI: 10.21203/rs.3.rs-496890/v1)

Date of submission: 9/8/2021

Abstract

Objective: To describe the development of a platform for image collection and annotation that resulted in a multi-sourced international image dataset of oral lesions to facilitate the development of automated lesion classification algorithms.

Materials and Methods: We developed a web-interface, hosted on a web server to collect oral lesions images from international partners. Further, we developed a customised annotation tool, also a web-interface for systematic annotation of images to build a rich clinically-labelled dataset. We evaluated the sensitivities comparing referral decisions through the annotation process with the clinical diagnosis of the lesions.

Results: The image repository hosts 2474 images of oral lesions consisting of oral cancer, oral potentially malignant disorders, and other oral lesions that were collected through MeMoSA® UPLOAD. Eight-hundred images were annotated by seven oral medicine specialists on MeMoSA®ANNOTATE, to mark the lesion and to collect clinical labels. The sensitivity in referral decision for all lesions that required a referral for cancer management/surveillance was moderate to high depending on the type of lesion (64.3 – 100%).

Conclusion: This is the first description of a database with clinically-labelled oral lesions. This database could accelerate the improvement of AI algorithms that can promote the early detection of high-risk oral lesions.

Introduction

Oral cancer (OC) was diagnosed in 377,713 individuals globally, with over 177,000 deaths in 2020 (Sung et al., 2020). OC disproportionately affects low- and middle-income countries (LMICs), and the majority are detected late resulting in poor survival (Warnakulasuriya, 2009). OC is often preceded by oral potentially malignant disorders (OPMD), which affords the opportunity to detect these lesions for management before the development of OC. Early detection of OPMD and OC requires trained healthcare practitioners who can differentiate high-risk lesions that are malignant or potentially malignant from those that do not have any risk of malignant transformation, thus enabling the appropriate management of oral lesions (Güneri & Epstein, 2014). Indeed, a lack of dental specialists has been associated with increased rates of delay in the detection of oral cancer (Crossman et al., 2016; Onizawa et al., 2003).

Applications of artificial intelligence (AI) techniques for disease detection, prognostication and prediction by medical image analysis are now being embraced for healthcare decision-making in medicine and dentistry (Joda, Yeung, Hung, Zitzmann, & Bornstein, 2020; Lindsell, Stead, & Johnson, 2020). A study by Esteva et al. has successfully used a form of AI called deep convolutional neural networks (CNNs) to classify skin lesions, identifying those that are malignant and most deadly, demonstrating that AI algorithms could reach a level of competence comparable to trained experts (Esteva et al., 2017). Such automated classification systems could be applied to

address the limited number of dental specialists that is often a bottleneck for clinical diagnosis in LMICs. The development of AI to automate the identification of OPMD and OC through clinical images is still in its infancy, mainly due to small image datasets and the lack of systematically annotated images of oral lesions (Camalan et al., 2021) that are required to develop such systems. Large training datasets with their corresponding ground truth (GT) and annotated labels are necessary to achieve good performance through efficient learning and to prevent overfitting (Krig, 2014; Mendonça, Ferreira, Marques, Marcal, & Rozeira, 2013; Yamashita, Nishio, Do, & Togashi, 2018).

We recognise the gap in this domain is a publicly accessible dataset and have initiated the collection of oral lesion images through selected members of the recently established Asia-Pacific Oral Cancer Network (APOCNET) (Syed Mohd Sobri, Kanapathy, Liew, & Cheong, 2020) and other clinical collaborators. The aim of the study is to report on the development of a platform that has enabled secure international transfer of images to the repository and a customized annotation tool for a uniform collection of clinically relevant information of images of the oral cavity that have hitherto not been available. Further, we performed GT analysis to identify lesions that are challenging to diagnose from images alone to improve the accuracy of annotations. The idea and motivation behind these platforms are to use the repository of images to develop an AI system for the automated classification of oral lesions based on the risk of malignant transformation (high risk vs low risk) (Lim et al., 2021; Welikala et al., 2020a, 2020b). When incorporated into a mobile phone application, such as MeMoSA® (Haron et al., 2021; Haron et al., 2020), a well-developed AI algorithm can facilitate early detection of OPMD and OC at the point-of-care.

Materials and Methods

Ethical approvals were obtained from the Institutional Review Board (IRB) of each centre for the use of anonymised images. To build a repository of well-annotated images, a platform to collect, store and annotate images securely was developed (Figure 1). Two components make up the platform; MeMoSA® UPLOAD and MeMoSA® ANNOTATE, which are linked to the MeMoSA® Data Vault (an image repository) through individual workbenches. All components are located on a secure cloud server. Authorised users with unique passwords can access the secure server via an encrypted Secure Socket Layer (SSL)/HTTPS connection. The server is maintained by an in-house system administrator.

MeMoSA® UPLOAD

MeMoSA® UPLOAD is a customised web interface, hosted on a web server used by clinical contributors to transfer images securely to the MeMoSA® Data Vault. Collaborators access their accounts to upload images with the associated metadata, including patient demographics and risk habits for each corresponding image. Uploaded data is transferred via secure network communication and stored in specific folders on the cloud server.

MeMoSA® ANNOTATE

MeMoSA® ANNOTATE is a customised annotation tool built on top of the open-source tool, ImageTagger (Fiedler, Bestmann, & Hendrich, 2018). MeMoSA® ANNOTATE was used for systematic annotation of the images in the MeMoSA® Data Vault. The images were chosen from the repository by a research team member to include a variety of lesions and of normal mucosa, and uploaded to MeMoSA® ANNOTATE with its metadata. The annotators, who are oral medicine specialists performed the annotation on their workbenches by accessing the network via an encrypted (SSL)/HTTPS connection.

MeMoSA® ANNOTATE enables the systematic annotation of images with labels describing the appearance of the lesion and captures referral recommendations made by the specialists. Guided by three board-certified oral medicine specialists, a decision tree that describes the flow of the tool was first developed. Seven main types of lesions were identified as significant clinical descriptors to describe an oral lesion: ulcer; white lesion, red lesion, mixed white and red lesion, swelling, pigmented lesion, and erosion (Scully, 2012). Lesions that did not fit into these descriptions such as oral submucous fibrosis (OSF) were categorized as "not applicable" and the lesion is named accordingly. Descriptions of the appearances including site, colour, presence of ulceration or swelling, texture, number of lesions, borders, and shape (Scully, 2012) were incorporated into the tool under the seven main types of lesions. Finally, the annotation process culminated in a referral decision and a classification of the lesion into a disease type (Table 1). Data on the image quality of each image was also collected.

Four board-certified oral medicine specialists validated clinical descriptors of the annotation tool by using a test image dataset consisting of eight OPMD, ten non-OPMD and two oral cancers

from Open Access sources, as there are no publicly available oral lesion image datasets. Open Access images were searched systematically using specific keywords “oral cancer”, “OPMD”, “benign lesions”, “developmental anomalies” and “normal anatomic variants” through Google Images. The images returned from the search were downloaded using Download All® Version 2.0.4. Each image was assessed and removed if they met the exclusion criteria as described in Supplementary Figure 1. Duplicate images were also removed. A questionnaire was administered to collect feedback on the accuracy of clinical descriptors, the feasibility of the annotation process and the user interface experience. The tool was modified on the basis of this feedback. The tool was further tested for accuracy of clinical descriptors by seven board-certified oral medicine specialists using 400 Open Access oral lesion images where each image was annotated by at least three specialists. These images included a variety of oral lesions and normal variants that fall into the following broad categories of OC, OPMD, non-OPMDs (benign lesions and developmental abnormalities (DA)) and normal mucosa/normal anatomical variants (NAV). A virtual workshop was held among all the annotators to reach a consensus on the descriptors that most define the respective lesions. Annotations included bounding boxes and labels, such as size, margin, site, outline were captured for each image in SQL format and stored in the PostgreSQL database in the MeMoSA® Data Vault. Once consensus was reached on the annotation labels, the annotation process continued with another set of 400 images with GT (61 images of OC, 90 of OPMD, 159 of benign lesions, 20 of DA and 70 of normal mucosa/NAV), where the images were annotated by seven board-certified oral medicine specialists.

The web server provides an application programming interface (API) for secure outward transfer of information from the MeMoSA® Data Vault (Figure 1). For example, to facilitate AI algorithm training using the annotated images, between members of our group, data can be downloaded as described above. The server is password protected and only accessible to an authorized user. The use of these images in the development of AI for the automated detection and classification of oral lesions as well as the development of a mouth landmark guidance tool has been published (Welikala et al., 2020a, 2020b, Lim et al., 2021).

Sensitivity of Annotations to Ground Truth

In addition to validating MeMoSA® ANNOTATE, we also set out to identify lesions that were difficult to diagnose and make referral decisions from visual images alone which could indicate

similar difficulties for the AI algorithm to classify lesions. To do this, we calculated the sensitivity of referral recommendations made by the annotators through MeMoSA® ANNOTATE compared to the clinical diagnosis (ground truth; GT), which is our gold standard for this study. We also computed accuracy, positive predictive value (PPV) and F1 scores for each referral category compared to GT. For this analysis, 400 images that were annotated on MeMoSA® ANNOTATE were analysed, consisting of 61 images of OC, 90 of OPMD, 159 of benign lesions, 20 of DA and 70 of normal mucosa/NAV.

OC and OPMD are classified as referable lesions, benign lesions and DA could be either 'refer for other reasons' or 'no referral' depending on the clinical attention required, while mucosa with NAV and those that do not have any changes would not require any referral and therefore categorized as 'no referral'. Sensitivity between the referral decision from the data collected through the annotation tool versus the GT was calculated for each disease category (OC, OPMD, benign, DA and normal mucosa/NAV) and for individual disease types for the OPMD category. Statistical analysis was conducted using SAS V9.4 (SAS Institutes, Cary, NC, USA).

Results

MeMoSA® UPLOAD

Two thousand seven hundred and three (2703) images were uploaded using MeMoSA® UPLOAD between June 2019 to September 2020, comprising of a variety of oral lesions as detailed in Table 1. The largest number of images collected was of benign lesions with 1041 images. This was followed by 539 images representing oral lichen planus (OLP), 482 images OSF and 298 images of OC. Images were assigned a unique sequential ID by the clinical collaborators as they were submitted through MeMoSA® UPLOAD. Manual assessment of the quality of uploaded images was done to i) determine image quality ii) identify images with potentially identifiable content and iii) remove duplicated images. Of the 2703 images, 105 images were excluded because they were out of focus, 105 images had extraoral content that could potentially identify a subject and another 19 images were excluded because they were duplicates. A subset of these images was randomly chosen to be annotated as described below.

MeMoSA® ANNOTATE

Images to be annotated are preloaded by the administrator and displayed as shown in Figure 2a. A selected image will appear along with its corresponding metadata (Figures 2b & c). Annotators will first annotate if a lesion is present on the image (Figure 2d). If present, a bounding box is placed around the lesion using their cursor, marking the location of the lesion (Figure 2e). The lesion type and its associated clinical descriptors to describe the appearance of the lesion is determined (Figure 2d) and a referral decision is decided (Figure 2f), followed by naming the lesion (Figure 2g). Finally, the image quality is rated (Figure 2h) before the annotations are saved (Figure 2i). Once submitted, the annotation is filed (Figure 2j). More than 800 images were annotated by seven oral medicine specialists.

Sensitivity of Annotations to Ground Truth

Of the 800 images, 400 images with GT annotated by seven specialists that resulted in 2800 annotations were included in our analysis. Overall, the sensitivities in the referral decision for all lesions that required a referral for cancer management or surveillance was moderate to high, ranging from 64.3% for erythroplakia and OSF up to 100% for discoid lupus erythematosus (DLE). For the lesions falling into the category of "refer-high risk" the overall sensitivity between the annotated decision and the GT to refer a lesion was 86.7%, and for those with cancer it was 90.6%. In this category, the lowest sensitivity was for erythroplakia and OSF at 64.3% for each of these categories. About 10.4% of the "refer-high risk" lesions were annotated to be "refer-low risk" where the majority of these fell into the non-homogenous leukoplakia disease type (30.2%) (Table 2). For lesions in the "refer-low risk" category, the overall sensitivity to refer a lesion was 85.7% with DLE having the highest sensitivity at 100% and homogenous leukoplakia at 80.2% being the lowest. Lesions in the "refer-low risk" category that were annotated as "refer-high risk" were mainly homogenous leukoplakia (16.5%). Overall, 13.3% of "refer-high risk" and 14.2% of "refer-low risk" lesions, were annotated as "refer for other reasons" or "no referral needed". The majority of these were OSF, erythroplakia and homogenous leukoplakia. Accuracy of 90.5% and 92.0% were achieved for "refer-high risk" and "refer-low risk" categories respectively. The PPV was 80.7% for "refer-high risk" and 71.5% for "refer-low risk", with F1 scores of 78.5% and 74.2% respectively (Supplementary Table 1). Benign lesions and developmental abnormalities were also correctly referred for other reasons/no referral with an accuracy of 88.8% and 93.6% respectively. Similarly, normal mucosa or mucosa with NAV were correctly identified as requiring no referral with a sensitivity of 83.7%, with only 5.7% annotated for referrals associated with

cancer management or surveillance (Table 2). Accuracy of 88.6% and 96.0% were achieved for "refer-other reasons" and "no referral" categories respectively. The positive predictive value was 85.8% for "refer-other reasons" and 93.0% for "no referral", with F1 scores of 87.5% and 88.1% respectively (Supplementary Table 1).

Discussion

Whilst emerging evidence demonstrates that AI could be used in classifying oral lesions (Camalan et al., 2021; Fu et al., 2020; Uthoff et al., 2018), progress has been slow due to the lack of a clinically labelled, well-annotated training dataset on oral lesions. MeMoSA® UPLOAD provides a standardised system to collect and uniformly transfer images and data, facilitating collaborations in many countries where the incidence of oral cancer is high. The large number of OC images collected in a short time represents the high burden of OC in these countries (Sung et al., 2020). The images collected were representative of the most prevalent oral mucosal lesions in South and South-East Asia, where the images originated. These included OLP, which is prevalent in Malaysia (Zain et al., 1997), while OSF is prevalent in Sri Lanka and Nepal (Amarasinghe, Johnson, Lalloo, Kumaraarachchi, & Warnakulasuriya, 2010; K. Warnakulasuriya et al., 1984). Similar efforts to establish image datasets have been seen in dermatology, in the HAM10000 (10,015 images) (Tschandl, Rosendahl, & Kittler, 2018) and PH2 (200 images) databases (Mendonça et al., 2013). Both these databases have facilitated machine learning for automated classification of skin lesions and accelerated the development in this field (Esteva et al., 2017). The Age-Related Eye Disease Study Research Group (2001) have also collected over 130,000 colour fundus images from 4,613 patients as a result of a 12-year longitudinal study for a better understanding of disease progression and risk factors behind macular degeneration. We believe that a large global network would accelerate the collection of a substantial number of diverse images of oral lesions that will cover all important diagnosis of oral lesions. MeMoSA® UPLOAD will allow the expansion to other clinical collaborators with ease to continue building on the database.

MeMoSA® ANNOTATE enables images to be annotated in a way that mimics the observations that a clinical specialist makes when examining a lesion. Tools similar to MeMoSA®ANNOTATE have been described in the literature. 'DerMat', used to annotate images in the PH2 database (Mendonça et al., 2013) allows users to draw and focus on a region of interest (Ferreira,

Mendonça, Rozeira, & Rocha, 2012). In the work of the Age-Related Eye Disease Study Research Group (2001), colour fundus images were graded by-hand by technicians and were used in several studies to train and validate deep learning algorithms (Burlina, Pacheco, Joshi, Freund, & Bressler, 2017; Burlina, Joshi, Pacheco, Liu, & Bressler, 2019; Burlina et al., 2017). The difference between MeMoSA®ANNOTATE and the tools mentioned is that it was built with a decision tree to capture or mimic a clinical oral examination by a specialist, and was customised to annotate features of oral lesions that cannot be done using the current tools for eye and skin diseases. Specialists manually annotated the images and referred to the metadata to arrive at a referral recommendation based on the risk of malignant transformation of the suspected lesion. Bounding boxes and clinical descriptions were collected to build a rich set of labels for each lesion. Further, the detailed annotated descriptions could improve AI through multiheaded training. Besides, this information could be used in the future to generate text rich reports for each referral recommendation by the AI. This could give clinicians more confidence in the type of information that has been used by the AI to make a referral decision. This database containing these annotated images has been used to guide the training of a deep learning algorithm, as described in recent publications (Welikala et al., 2020a, 2020b).

We observed that sensitivity was high in providing the correct referral decision for cancer surveillance compared to the GT, with more than 85% for all disease types except erythroplakia and OSF. Regarding DLE, we report 100% agreement in the referral decision, although this is a lesion that is often difficult to distinguish from OLP (Warnakulasuriya, 2018). However, there was only one image of DLE in our 400-image cohort and all seven annotators correctly identified this as a referrable low-risk lesion (Table 1). The high sensitivity shows that the data input in the form of images, and information on patient demographics and risk habits are reliable in making a referral decision and should be reliable information for AI training. However, further refinements can be made. Some lesions have been identified to be difficult to make decisions based on images alone, such as early OSF, which is particularly difficult to diagnose even during a clinical oral examination, as it can appear as blanching of the oral mucosa or loss of pigmentation which only appears in some South Asian populations (Warnakulasuriya, 2018). Clinical diagnosis is usually reached after examination of mucosa and evaluating clinical history and information on symptoms such as burning sensation, dry mouth or limited mouth opening along with the presence of palpable fibrous bands. Therefore, such information should be collected as metadata in

anticipation that these would also be useful in the training of the AI algorithm. As for erythroplakia, 35.7% were annotated as "refer for other reasons", by identifying it as a benign lesion. Clinically, erythroplakia could be mistaken for several red-appearing benign conditions, including erythematous candidiasis or erythema migrans (Warnakulasuriya, 2018). However, the review of a single image hampered the ability of the specialist to distinguish erythroplakia from these benign lesions. We will need to consider displaying multiple images of the oral cavity for each patient to present a comprehensive view to obtain more accurate annotations or collect clinical information and medical history based on questions that enable the capture of these information. As for lesions that did not require a referral, the high positive predictive values and F1 scores for "refer for other reasons" and "no referral needed" categories indicate these lesions were correctly identified and not referred, meaning the AI could reduce the burden to the healthcare system due to inaccurate referral should it be implemented.

Limitations

Whilst patient demographics and risk habit information are available, further questions to collect more comprehensive clinical information that could help annotators reach a more accurate decision should be considered. This is particularly true for the OPMD lesions such as erythroplakia and leukoplakia which are by definition diagnosed based on the exclusion of other diagnoses. Therefore, in the next developmental stage of MeMoSA®ANNOTATE, avenues for richer data collection possibly including patient history such as chief complaint, history of chief complaint, medical history and oral hygiene products used would be incorporated. We had to remove images that were out of focus or have personal identifiers and to mitigate this, we developed an image capturing protocol to standardize image capture. We are also developing an AI-assisted mouth landmark guidance tool that will aid the user in the image capture process (Lim et al., 2021). In addition, the image collection is only done in Asia and there is ongoing work to establish a global network to expand the use of the tool to collect a larger number of images from various oral lesions.

Conclusion

We described the establishment of a systematic and secure platform for the development of an oral lesion image repository that has hitherto not been available. Given that clinical oral examination can be conducted relatively easily by a primary healthcare practitioner, automated detection and

classification algorithm would go a long way in helping clinicians distinguish the multitude of types of oral lesions and NAV that occur in the oral cavity. Our work as described here could accelerate the progress in developing such an AI system.

Acknowledgements

We would like to thank Prof Mahima V Guledgud of JSS Dental College & Hospital and Dr Anand Ramanathan of the University of Malaya for initial discussions on the project. The authors are grateful to Dr Felix Anthony Wiranata Wong of Trisakti University, Indonesia for technical assistance in collecting images. We thank the supporters of Cancer Research Malaysia. Cancer Research Malaysia is a non-profit research organization. We are committed to understanding cancer prevention, diagnosis, and treatment through a fundamental research program.

Disclosure Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding source

This work was supported by the Medical Research Council [grant number MR/S013865/1]. The funders had no involvement in the conduct of the research or preparation of the manuscript.

Authors' Contributions

Senthilmani Rajendran contributed to design, data acquisition, analysis and interpretation, drafted and critically revised the manuscript; Jian Han Lim, Kohgulakuhan Yogalingam and Lim Ying Zhi contributed to data acquisition and drafted manuscript; Thomas George Kallarakkal, Ruwan Duminda Jayasinghe, Jyotsna Rimal, Rahmi Amtha, Karthikeya Patil and Roshan Alex Welikala contributed to data acquisition and critically revised the manuscript; Rosnah Binti Zain contributed to conception and design, data acquisition and critically revised the manuscript; Alexander Ross Kerr contributed to design, data acquisition and critically revised the manuscript; Paolo Remagnino, John Gibson, Wanninayake Mudiyanselage Tilakaratne, Sarah Ann Barman and Chee Seng Chan contributed to conception and critically revised the manuscript; Chee Sun Liew contributed to design, data acquisition and critically revised the manuscript; Yi-Hsin Yang contributed to data analysis and interpretation, and critically revised the manuscript; Sok Ching

Cheong contributed to conception and design, data acquisition, analysis and interpretation, drafted and critically revised the manuscript. All authors gave their final approval and agree to be accountable for all aspects of the work.

Ethics Approval

This study was performed in line with the principles of the Declaration of Helsinki. Ethical approval for the transfer and use of images for this study was granted from all provider institutions in Sri Lanka (ERC/FDS/UOP/I/2019/4), Indonesia (249/S3/KEPK/FKG/3/2019), Malaysia (UM ethics – DF OS1902/0011(L)/1970) and Nepal (ref no. 1430) and Kingston University (project no. 083.1).

Consent to participate

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

Consent to publication

No identifying information was included in this article. However, all participants provided informed consent to publish their anonymised images.

Data Sharing and Data Accessibility

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

Amarasinghe, H., Johnson, N., Lalloo, R., Kumaraarachchi, M., & Warnakulasuriya, S. (2010). Derivation and validation of a risk-factor model for detection of oral potentially malignant disorders in populations with high prevalence. *British journal of cancer, 103*(3), 303-309. doi:10.1038/sj.bjc.6605778

Burlina, P., Pacheco, K. D., Joshi, N., Freund, D. E., & Bressler, N. M. (2017). Comparing humans and deep learning performance for grading AMD: a study in using universal deep features and transfer learning for automated AMD analysis. *Computers in biology and medicine, 82*, 80-86. doi:10.1016/j.compbiomed.2017.01.018

Burlina, P. M., Joshi, N., Pacheco, K. D., Liu, T. A., & Bressler, N. M. (2019). Assessment of deep generative models for high-resolution synthetic retinal image generation of age-related macular degeneration. *JAMA ophthalmology, 137*(3), 258-264. doi:10.1001/jamaophthalmol.2018.6156

Burlina, P. M., Joshi, N., Pekala, M., Pacheco, K. D., Freund, D. E., & Bressler, N. M. (2017). Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. *JAMA ophthalmology, 135*(11), 1170-1176. doi:10.1001/jamaophthalmol.2017.3782

Camalan, S., Mahmood, H., Binol, H., Araújo, A. L. D., Santos-Silva, A. R., Vargas, P. A., . . . Gurcan, M. N. (2021). Convolutional Neural Network-Based Clinical Predictors of Oral Dysplasia: Class Activation Map Analysis of Deep Learning Results. *Cancers, 13*(6), 1291. doi:https://doi.org/10.3390/cancers13061291

Crossman, T., Warburton, F., Richards, M. A., Smith, H., Ramirez, A., & Forbes, L. J. (2016). Role of general practice in the diagnosis of oral cancer. *British Journal of Oral and Maxillofacial Surgery, 54*(2), 208-212. doi:https://doi.org/10.1016/j.bjoms.2015.11.003

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *nature, 542*(7639), 115-118. doi:https://doi.org/10.1038/nature21056

Ferreira, P. M., Mendonça, T., Rozeira, J., & Rocha, P. (2012). *An annotation tool for dermoscopic image segmentation.* Paper presented at the Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications.

Fiedler, N., Bestmann, M., & Hendrich, N. (2018). *Imagetagger: An open source online platform for collaborative image labeling.* Paper presented at the Robot World Cup.

Fu, Q., Chen, Y., Li, Z., Jing, Q., Hu, C., Liu, H., . . . Li, K. (2020). A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: A retrospective study. *EClinicalMedicine, 27*, 100558.

Güneri, P., & Epstein, J. B. (2014). Late stage diagnosis of oral cancer: components and possible solutions. *Oral oncology, 50*(12), 1131-1136. doi:10.1016/j.oraloncology.2014.09.005

Haron, N., Rajendran, S., Kallarakkal, T. G., Zain, R. B., Ramanathan, A., Abraham, M. T., . . . Cheong, S. C. (2021). High Referral Accuracy for Oral Cancers and Oral Potentially

Malignant Disorders using Telemedicine. *Oral diseases*. doi:https://doi.org/10.1111/odi.13892

Haron, N., Zain, R. B., Ramanathan, A., Abraham, M. T., Liew, C. S., Ng, K. G., . . . Thangavalu, L. A. P. (2020). m-Health for Early Detection of Oral Cancer in Low-and Middle-Income Countries. *Telemedicine and e-Health, 26*(3), 278-285. doi:10.1089/tmj.2018.0285

Joda, T., Yeung, A., Hung, K., Zitzmann, N., & Bornstein, M. (2020). Disruptive Innovation in Dentistry: What It Is and What Could Be Next. *Journal of dental research, 16*, 22034520978774. doi:10.1177/0022034520978774

Krig, S. (2014). Ground Truth Data, Content, Metrics, and Analysis. In *Computer Vision Metrics: Survey, Taxonomy, and Analysis* (pp. 283-311). Berkeley, CA: Apress.

Lim, J. H., Tan, C. S., Chan, C. S., Welikala, R. A., Remagnino, P., Rajendran, S., . . . Barman, S. A. (2021). *D'OraCa: Deep Learning-Based Classification of Oral Lesions with Mouth Landmark Guidance for Early Detection of Oral Cancer.* Paper presented at the Annual Conference on Medical Image Understanding and Analysis.

Lindsell, C. J., Stead, W. W., & Johnson, K. B. (2020). Action-Informed Artificial Intelligence—Matching the Algorithm to the Problem. *Jama, 323*(21), 2141-2142. doi:10.1001/jama.2020.5035

Mendonça, T., Ferreira, P. M., Marques, J. S., Marcal, A. R., & Rozeira, J. (2013). *PH 2-A dermoscopic image database for research and benchmarking.* Paper presented at the 2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC).

Onizawa, K., Nishihara, K., Yamagata, K., Yusa, H., Yanagawa, T., & Yoshida, H. (2003). Factors associated with diagnostic delay of oral squamous cell carcinoma. *Oral oncology, 39*(8), 781-788. doi:10.1016/s1368-8375(03)00075-7

Scully, C. (2012). *Oral and Maxillofacial Medicine-E-Book: The Basis of Diagnosis and Treatment*: Elsevier Health Sciences.

Sung, H., Ferlay, J., Siegel, R., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2020). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin, 2021*(71), 209-249. doi:https://doi.org/10.3322/caac.21660

Syed Mohd Sobri, S. N. S., Kanapathy, J., Liew, C. S., & Cheong, S. C. (2020). The Establishment of the Asia-Pacific Oral Cancer Network (APOCNET)-Inaugural Stakeholders' Meeting. *Oral diseases*. doi:10.1111/odi.13320

Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data, 5*, 180161. doi:https://doi.org/10.1038/sdata.2018.161

Uthoff, R. D., Song, B., Sunny, S., Patrick, S., Suresh, A., Kolur, T., . . . Wilder-Smith, P. (2018). Point-of-care, smartphone-based, dual-modality, dual-view, oral cancer screening device with neural network classification for low-resource communities. *PLOS ONE, 13*(12), e0207493. doi:https://doi.org/10.1371/journal.pone.0207493

Warnakulasuriya, K., Ekanayake, A., Sivayoham, S., Stjernswärd, J., Pindborg, J., Sobin, L., & Perera, K. (1984). Utilization of primary health care workers for early detection of oral cancer and precancer cases in Sri Lanka. *Bulletin of the World Health Organization, 62*(2), 243.

Warnakulasuriya, S. (2009). Global epidemiology of oral and oropharyngeal cancer. *Oral oncology, 45*(4-5), 309-316. doi:0.1016/j.oraloncology.2008.06.002

Warnakulasuriya, S. (2018). Clinical features and presentation of oral potentially malignant disorders. *Oral surgery, oral medicine, oral pathology and oral radiology, 125*(6), 582-590. doi:10.1016/j.oooo.2018.03.011

Welikala, R. A., Remagnino, P., Lim, J. H., Chan, C. S., Rajendran, S., Kallarakkal, T. G., . . . Barman, S. A. (2020a). Automated detection and classification of oral lesions using deep learning for early detection of oral cancer. *IEEE Access, 8*, 132677-132693. doi:10.1109/ACCESS.2020.3010180

Welikala, R. A., Remagnino, P., Lim, J. H., Chan, C. S., Rajendran, S., Kallarakkal, T. G., . . . Barman, S. A. (2020b). *Fine-Tuning Deep Learning Architectures for Early Detection of Oral Cancer.* Paper presented at the Mathematical and Computational Oncology: Second International Symposium, ISMCO 2020, San Diego, CA, USA, October 8–10, 2020, Proceedings.

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into imaging, 9*(4), 611-629. doi:https://doi.org/10.1007/s13244-018-0639-9

Zain, R. B., Ikeda, N., Razak, I. A., Axéll, T., Majid, Z. A., Gupta, P. C., & Yaacob, M. (1997). A national epidemiological survey of oral mucosal lesions in Malaysia. *Community dentistry and oral epidemiology, 25*(5), 377-383. doi:10.1111/j.1600-0528.1997.tb00959.x

Figure Legends

Figure 1 Platform for secure image transfer, collection, storage and annotation. (a) MeMoSA® UPLOAD to securely transfer images; (b) MeMoSA® ANNOTATE for clinical annotations of selected images by specialists; (c) secure outward transfer of information from the MeMoSA® Data Vault for use in AI training

Figure 2 MeMoSA® ANNOTATE components. (a) Anonymised images to be annotated; (b) Image selected for annotation; (c) Metadata associated with selected image; (d) Annotating the presence of a lesion and the features of the lesion (e) Bounding box to mark the lesion; (f) Referral decision (g) Lesion type; (h) Image quality; (i) Save annotations; (j) Saved annotation displayed

Table 1 List of referable and non-referrable lesions and the number of images collected through MeMoSA® UPLOAD

Disease Type	**Number of images**
Referrable lesions – cancer/high risk OPMD	
Oral Cancer/Suspicious of Oral Cancer	298
Non-homogeneous leukoplakia	47
Erythroplakia	18
Oral submucous fibrosis	482
Verrucous Hyperplasia	2
Referrable lesions – low risk OPMD	
Homogeneous leukoplakia	90
Lichenoid lesion/Lichen planus	539
Discoid lupus erythematosus (other than lip)	29
Referrable lesions - for other reasons/no referral needed	
Benign	1041
Developmental abnormalities	23
No referral needed	
Normal anatomical variant	73
Normal mucosa	61
Total	**2703**

Table 2 Sensitivity of referral decisions made following annotations on MeMoSA® ANNOTATE compared to the Ground Truth

Referral Decision	Disease type based on ground truth	Sensitivity of annotations in MeMoSA®ANNOTATE compared to Ground Truth, % (n refers to number of annotations)									
		Refer[a] (%)	(n)	Refer-high risk (%)	(n)	Refer-low risk (%)	(n)	Refer-other reasons /No referral (%)	(n)	No referral (normal mucosa) (%)	(n)
Refer-high risk	**Overall (n = 637)**	86.7	552	**76.3**	**486**	10.4	66	11.6	74	1.7	11
	Oral Cancer (n = 427)	90.6	387	**87.1**	**372**	3.5	15	9.1	39	0.2	1
	Non-homogenous leukoplakia, NHL (n = 126)	88.1	111	**57.9**	**73**	30.2	38	11.9	15	0	0
	Erythroplakia (n = 14)	64.3	9	**57.1**	**8**	7.1	1	35.7	5	0	0
	Oral submucous fibrosis, OSF (n = 70)	64.3	45	**47.1**	**33**	17.1	12	21.4	15	14.3	10
Refer-low risk	**Overall (n = 420)**	85.7	360	8.6	36	**77.1**	**324**	14	59	0.2	1
	Homogenous leukoplakia, HL (n = 91)	80.2	73	16.5	15	**63.7**	**58**	18.7	17	1.1	1
	Lichenoid lesion / Oral lichen planus, OLP (n = 322)	87	280	6.5	21	**80.4**	**259**	13.1	42	0	0
	Discoid lupus erythematous, DLE (n = 7)	100	7	0	0	**100**	**7**	0	0	0	0
Refer-other	**Overall (n = 1253)**	9.2	115	5.6	70	3.6	45	**89.3**	**1119**	1.5	19

reasons / no referral needed	**Benign (n = 1113)**	10.1	112	6.3	70	3.8	42	**88.8**	**988**	1.2	13
	Developmental anomalies, DA (n = 140)	2.1	3	0	0	2.1	3	**93.6**	**131**	4.3	6
No referral (normal mucosa)	**Normal anatomical variant/no lesion (n = 490)**	5.7	28	2	10	3.7	18	10.6	52	**83.7**	**410**

[a]Refer: Both refer -high and -low risk

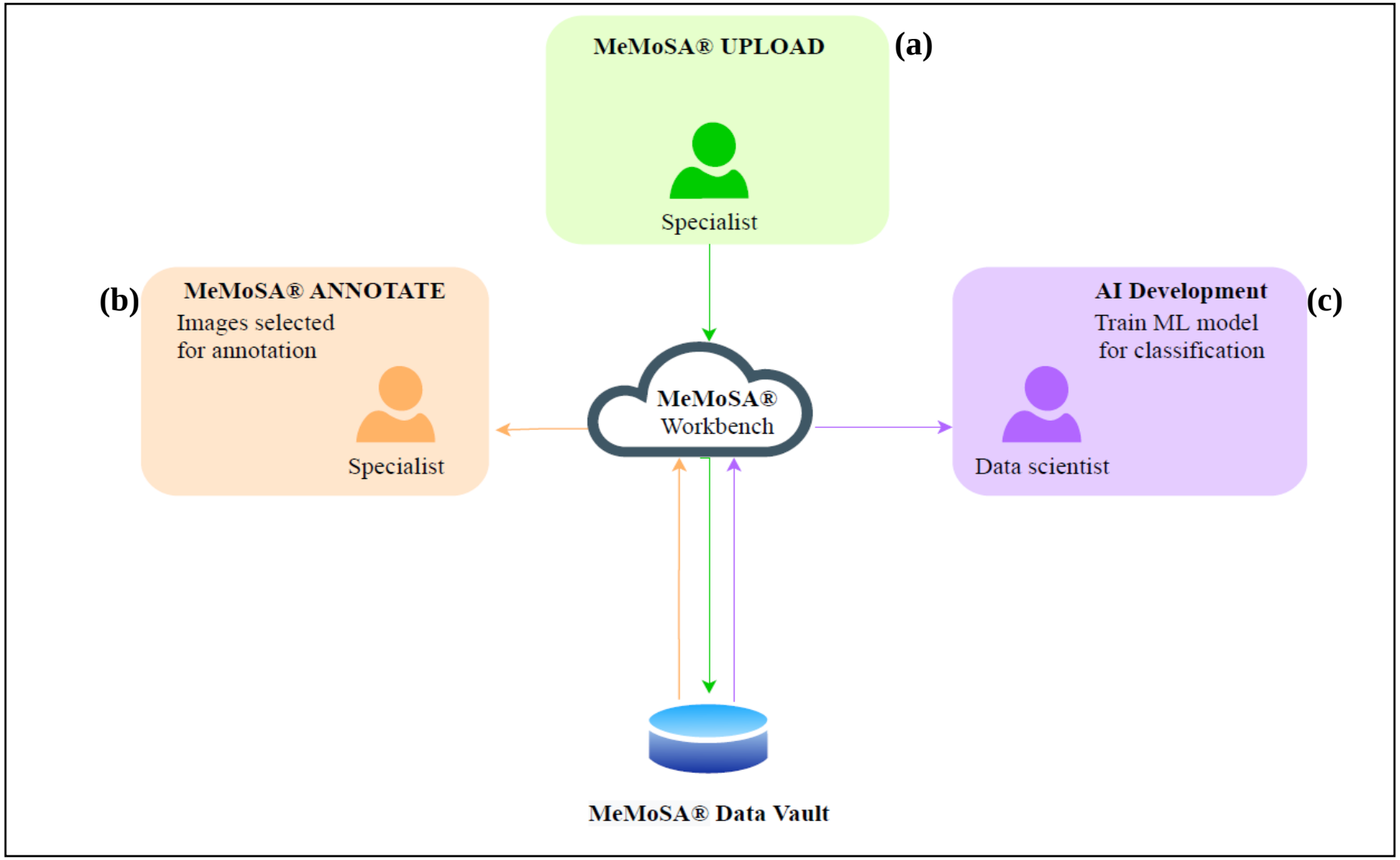

Figure 1 Platform for secure image transfer, collection, storage and annotation. (a) MeMoSA® UPLOAD to securely transfer images; (b) MeMoSA® ANNOTATE for clinical annotations of selected images by specialists; (c) secure outward transfer of information from the MeMoSA® Data Vault for use in AI training

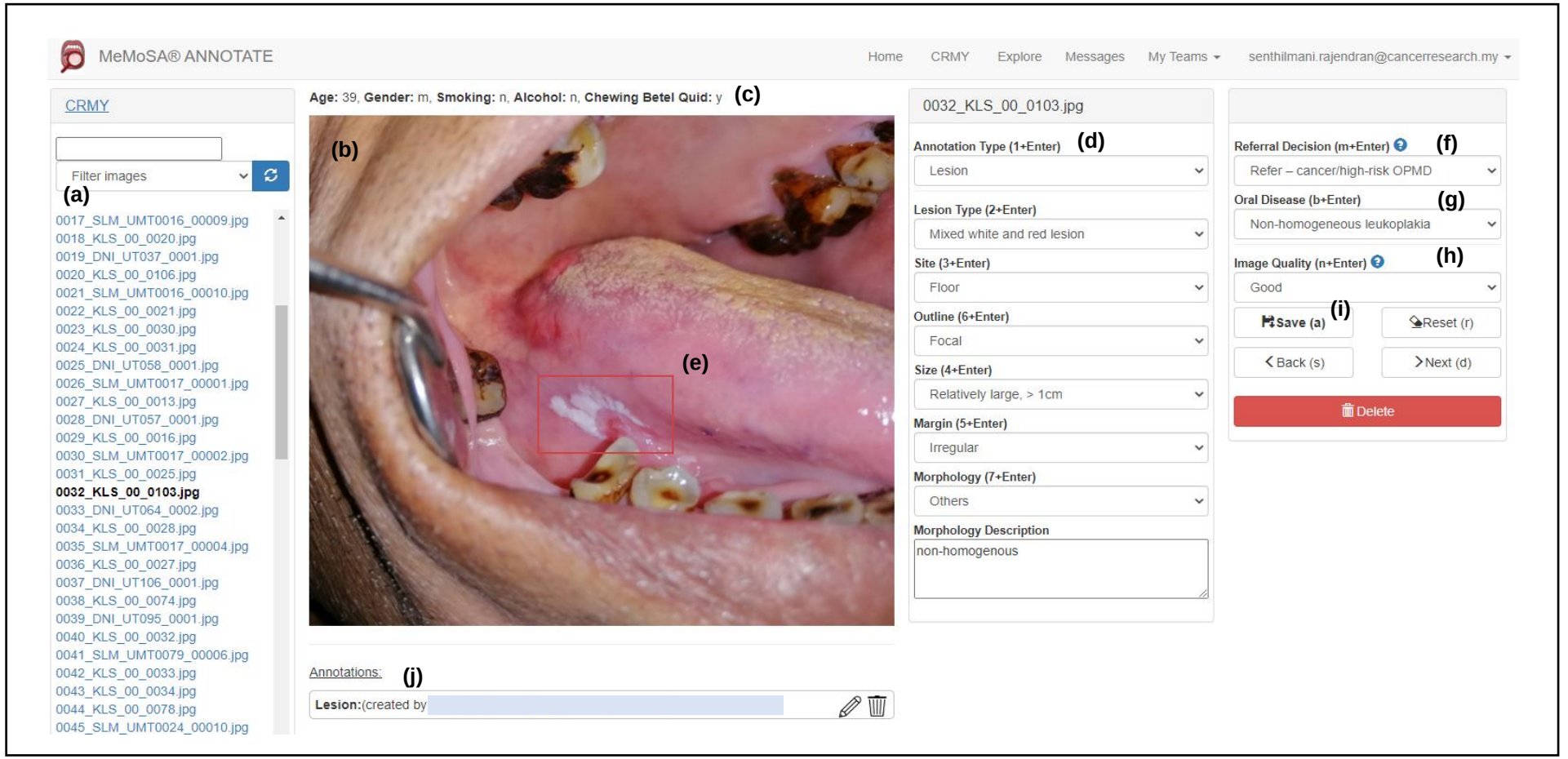

Figure 2 MeMoSA® ANNOTATE components. (a) Anonymised images to be annotated; (b) Image selected for annotation; (c) Metadata associated with selected image; (d) Annotating the presence of a lesion and the features of the lesion (e) Bounding box to mark the lesion; (f) Referral decision (g) Lesion type; (h) Image quality; (i) Save annotations; (j) Saved annotation displayed