ORIGINAL ARTICLE



## Addressing data imbalance challenges in oral cavity histopathological whole slide images with advanced deep learning techniques

 $\label{eq:assistance} Tabasum \ Majeed^1 \cdot Tariq \ Ahmad \ Masoodi^2 \cdot Muzafar \ Ahmad \ Macha^3 \cdot Muzafar \ Rasool \ Bhat^4 \cdot Khalid \ Muzaffar^5 \cdot Assif \ Assad^{1} \\ \end{tabular}$ 

Received: 18 November 2023 / Revised: 9 May 2024 / Accepted: 7 July 2024

© The Author(s) under exclusive licence to The Society for Reliability Engineering, Quality and Operations Management (SREQOM), India and The Division of Operation and Maintenance, Lulea University of Technology, Sweden 2024

Abstract Oral Cavity Squamous Cell Carcinoma (OCSCC) represents a common form of head and neck cancer originating from the mucosal lining of the oral cavity, often detected in advanced stages. Traditional detection methods rely on analyzing hematoxylin and eosin (H&E)stained histopathological whole-slide images, which are time-consuming and require expert pathology skills. Hence, automated analysis is urgently needed to expedite diagnosis and improve patient outcomes. Deep learning, through automated feature extraction, offers a promising avenue for capturing high-level abstract features with greater accuracy than traditional methods. However, the imbalance in class distribution within datasets significantly affects the performance of deep learning models during training, necessitating specialized approaches. To address the issue, various methods have been proposed at both data and algorithmic levels. This study investigates strategies to mitigate class imbalance by employing a publicly available OCSCC imbalance dataset. We evaluated undersampling methods (Near Miss, Edited

Assif Assad assif.assad@islamicuniversity.edu.in

Tabasum Majeed tabasum.majeed@iust.ac.in

Tariq Ahmad Masoodi tahamasoodi@gmail.com

Muzafar Ahmad Macha muzafar.macha@iust.ac.in

Muzafar Rasool Bhat muzafar.rasool@islamicuniversity.edu.in

Khalid Muzaffar khalid.muzaffar@islamicuniversity.edu.in

<sup>1</sup> Department of Computer Science and Engineering, Islamic University of Science and Technology, Awantipora, Kashmir, India Nearest Neighbors) and oversampling techniques (SMOTE, Deep SMOTE, ADASYN) integrated with transfer learning across different imbalance ratios (0.1, 0.15, 0.20, 0.30). Our findings demonstrate the effectiveness of SMOTE in improving test performance, highlighting the efficacy of strategic oversampling combined with transfer learning in classifying imbalanced medical datasets. This enhances OCSCC diagnostic accuracy, streamlines clinical decisions, and reduces reliance on costly histopathological tests.

**Keywords** Imbalanced data · Deep learning · Histopathology · Whole slide Images · Oral cavity squamous cell carcinoma · Oversampling · Undersampling

## 1 Introduction

Oral squamous cell carcinoma (OSCC), which is the predominant form of cancer in the head and neck region originates

- <sup>2</sup> Department of Cancer Research, Sidra Medicine, Doha, Qatar
- <sup>3</sup> Watson-Crick Centre for Molecular Medicine, Islamic University of Science and Technology, Awantipora, Kashmir, India
- <sup>4</sup> Department of computer science, Islamic University of Science and Technology, Awantipora, Kashmir, India
- <sup>5</sup> Department of Electronics and Communication Engineering, Islamic University of Science and Technology, Awantipora, Kashmir, India