JointViT: Modeling Oxygen Saturation Levels with Joint Supervision on Long-Tailed OCTA

Zeyu Zhang, Xuyin Qi, Mingxi Chen, Guangxi Li, Ryan Pham, Ayub Qassim, Ella Berry, Zhibin Liao, Owen Siggs, Robert Mclaughlin, Jamie Craig, Minh-Son To

Flinders University, The Australian National University, The University of Adelaide, Guangdong Technion - Israel Institute of Technology, The University of Sydney

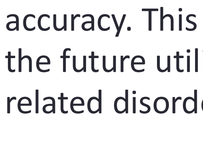
ABSTRACT

The oxygen saturation level in the blood (SaO2) is crucial for health, particularly in relation to sleep-related breathing. However, continuous monitoring of SaO2 is time-consuming and highly variable. Recently, optical coherence tomography angio-graphy (OCTA) has shown promising development in rapidly and effectively screening eye-related lesions, offering the poten tial for diagnosing sleep-related disorders. To bridge this gap, we propose JointViT, a novel model based on the ViT architecture, in-corporating a joint loss function for supervision and a balancing augmentation technique during data preprocessing to improve the model's performance.

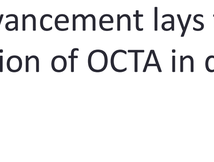
INTRODUCTION

Obstructive sleep apnoea (OSA) is the most common sleep-related disorder. Low blood oxygen levels during sleep may indicate the presence of OSA or a sleep-related breathing disorder. Oxygen saturation level in the blood (SaO_2) indicates the percentage of hemoglobin binding sites in the bloodstream occupied by oxygen molecules. This metric serves as a crucial indicator of the body's ability to adequately oxygenate tissues and organs. Conducting sleep studies is time-consuming and demands continuous monitoring. Optical coherence tomography (OCT) is a non-invasive imaging test that utilises infrared light to obtain 2D and 3D visualisation of structures with micrometre resolution. OCT angiography (OCTA) is an extension of the OCT technique that extracts blood flow information of the retina, which allows both vessel structure and haemodynamic parameters to be measured. The ability of OCTA to image the microvasculature of the retina opens the door to screening for sleep-related disordered . Since the OCTA data is acquired in hospitals where there are more patients with sleep-related disorders and fewer healthy instances, this leads to a long-tailed distribution, making the prediction challenging.

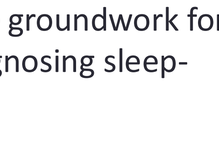
SaO_2 (%)	Classification
96 - 100	Normal
93 - 95	Borderline Low
89 - 92	Low



(a) Low

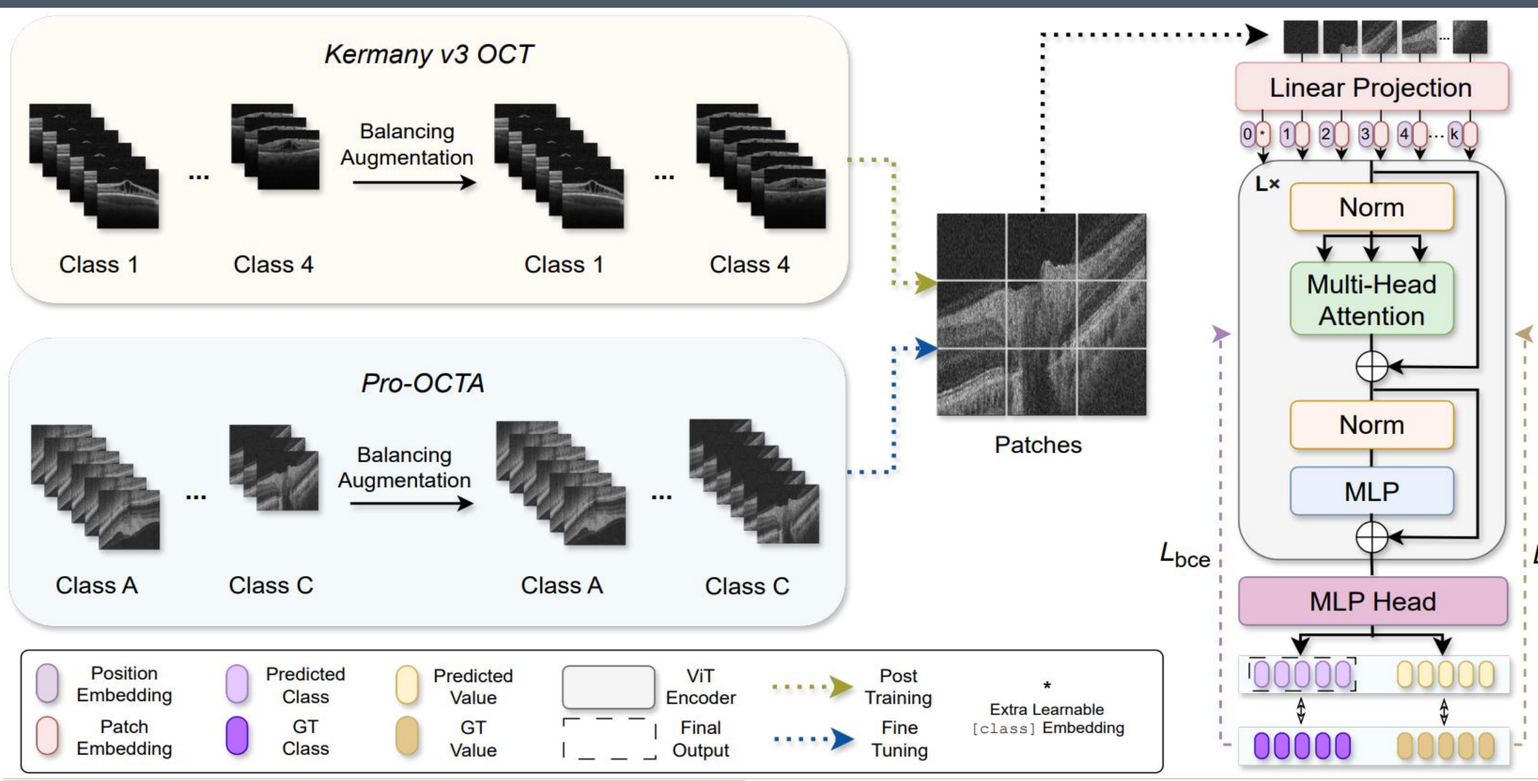


(b) Borderline Low



(c) Normal

We have proposed a novel approach grounded in ViT architecture to leverage OCTA data for predicting SaO_2 categories. Our contributions:
1. proposed a novel model named **JointViT** built upon the plain ViT architecture, which incorporates a well-designed joint loss function that leverages both categories and values for supervision.
2. introduced a balancing augmentation technique within the data pre-processing phase to enhance the model's performance specifically on the long-tail distribution within the OCTA dataset.3. conducted comprehensive experiments on the OCTA dataset, which demonstrated that our proposed method has significantly outperformed other state-of-the-art methods, yielding improvements of up to 12.28% in accuracy. This advancement lays the groundwork for the future utilization of OCTA in diagnosing sleep-related disorders.



METHODOLOGY

We employ both SaO_2 categories Y_{cls} and exact SaO_2 values Y_{val} as joint supervision.
For classification, we utilize binary cross-entropy loss L_{bce}. For regression, we adopt mean squared error loss L_{mse} for supervision. The weight of each loss in the joint loss is controlled by coefficient λ.
$L = \lambda L_{bce}(T(X), Y_{cls}) + (1- \lambda)L_{mse}(T(X), Y_{val})$

COMPARATIVE STUDY

We compared our proposed **JointViT** on Prog-OCTA dataset with well-established 2D and 3D methods which are widely used in medical imaging recognition. The results demonstrate that our method significantly outperforms others in predicting saturation levels using imbalanced OCTA data.

Models	Test Acc. ↑	Sensitivity ↑	Specificity ↑
M3T [14]	$61.40^{\pm 2.48}$	$33.33^{\pm 0.00}$	$66.67^{\pm 0.00}$
I3D [2]	$61.40^{\pm 2.48}$	$33.33^{\pm 0.00}$	$66.67^{\pm 0.00}$
MedViT [22]	$61.40^{\pm 2.48}$	$33.33^{\pm 0.00}$	$66.67^{\pm 0.00}$
FCNlinksCNN [28]	$57.89^{\pm 7.44}$	$31.31^{\pm 2.86}$	$64.96^{\pm 2.42}$
MedicalNet (3D ResNet-10) [4]	$63.16^{\pm 0.00}$	$37.96^{\pm 3.47}$	$68.95^{\pm 1.72}$
MedicalNet (3D ResNet-18) [4]	$63.16^{\pm 4.30}$	$40.74^{\pm 8.59}$	$71.62^{\pm 5.52}$
MedicalNet (3D ResNet-34) [4]	$61.40^{\pm 2.48}$	$33.33^{\pm 0.00}$	$66.67^{\pm 0.00}$
MedicalNet (3D ResNet-50) [4]	$61.40^{\pm 2.48}$	$33.33^{\pm 0.00}$	$66.67^{\pm 0.00}$
MedicalNet (3D ResNet-101) [4]	$61.40^{\pm 2.48}$	$33.33^{\pm 0.00}$	$66.67^{\pm 0.00}$
MedicalNet (3D ResNet-152) [4]	$61.40^{\pm 2.48}$	$33.33^{\pm 0.00}$	$66.67^{\pm 0.00}$
MedicalNet (3D ResNet-200) [4]	$61.40^{\pm 2.48}$	$33.33^{\pm 0.00}$	$66.67^{\pm 0.00}$
3D DenseNet-121 [13]	$66.67^{\pm 2.48}$	$45.37^{\pm 3.46}$	$72.58^{\pm 0.82}$
3D DenseNet-161 [13]	$66.67^{\pm 2.48}$	$47.22^{\pm 2.27}$	$75.06^{\pm 2.14}$
3D DenseNet-169 [13]	$66.67^{\pm 2.48}$	$43.70^{\pm 7.97}$	$72.90^{\pm 2.37}$
3D DenseNet-201 [13]	$64.91^{\pm 4.96}$	$39.81^{\pm 4.72}$	$71.47^{\pm 3.40}$
COVID-ViT [8]	$61.40^{\pm 6.40}$	$36.63^{\pm 7.76}$	$61.66^{\pm 8.13}$
JointViT (Ours)	$\mathbf{70.17}^{\pm 8.90}$	$\mathbf{62.51}^{\pm 8.44}$	$\mathbf{82.83}^{\pm 7.07}$

ABLATION STUDY

Ablation results in the table shows the impact of varying joint loss coefficient (λ) values on model performance, which indicates when λ is set to be 0.99, the model achieved optimal performance.

Models	Test Acc. ↑	Sensitivity ↑	Specificity ↑
COVID-ViT [8]	$61.40^{\pm 6.40}$	$36.63^{\pm 7.76}$	$61.66^{\pm 8.13}$
JointViT ($\lambda = 0.8$)	$61.40^{\pm 4.71}$	$33.35^{\pm 5.44}$	$67.41^{\pm 7.27}$
JointViT ($\lambda = 0.9$)	$65.72^{\pm 2.72}$	$48.89^{\pm 4.20}$	$66.27^{\pm 3.13}$
JointViT ($\lambda = 0.95$)	$64.91^{\pm 4.21}$	$54.67^{\pm 6.11}$	$70.31^{\pm 3.82}$
JointViT ($\lambda = 0.96$)	$63.61^{\pm 6.67}$	$50.43^{\pm 4.51}$	$76.20^{\pm 2.10}$
JointViT ($\lambda = 0.97$)	$66.67^{\pm 2.46}$	$57.22^{\pm 7.31}$	$64.82^{\pm 6.20}$
JointViT ($\lambda = 0.98$)	$66.67^{\pm 4.50}$	$51.32^{\pm 3.42}$	$70.42^{\pm 6.26}$
JointViT ($\lambda = 1$)	$63.16^{\pm 0.00}$	$35.92^{\pm 3.21}$	$76.81^{\pm 2.29}$
JointViT ($\lambda = 0.99$)	$\mathbf{70.17}^{\pm 8.90}$	$\mathbf{62.51}^{\pm 8.44}$	$\mathbf{82.83}^{\pm 7.07}$

The table presents the results of conducting ablations involving the integration of different loss functions within the joint loss framework. The results indicate that our jointly designed loss function, which combines BCE and MSE losses, consistently achieves superior performance compared to alternative configurations.

Models	Test Acc. ↑	Sensitivity ↑	Specificity ↑
JointViT w/o joint loss	$63.16^{\pm 0.00}$	$35.92^{\pm 3.21}$	$76.81^{\pm 2.29}$
JointViT w/ Bal-BCE ($\lambda = 1$)	$61.40^{\pm 2.48}$	$20.47^{\pm 0.82}$	$53.80^{\pm 0.83}$
JointViT w/ Bal-BCE ($\lambda = 0.99$)	$64.91^{\pm 2.48}$	$45.63^{\pm 17.55}$	$71.08^{\pm 11.96}$
JointViT w/ joint loss ($\lambda = 0.99$)	$\mathbf{70.17}^{\pm 8.90}$	$\mathbf{62.51}^{\pm 8.44}$	$\mathbf{82.83}^{\pm 7.07}$

We explored the efficacy of employing the second-best backbone . Our findings reveal that our approach which used a ViT backbone maintains its position as the top-performing method.

Models	Test Acc. ↑	Sensitivity ↑	Specificity ↑
2D DenseNet-161 w/ bal-aug & joint loss	$59.65^{\pm 4.96}$	$26.26^{\pm 7.36}$	$60.17^{\pm 8.17}$
3D DenseNet-161 w/ bal-aug & joint loss	$55.09^{\pm 3.46}$	$44.82^{\pm 6.89}$	$68.37^{\pm 10.46}$
JointViT	$\mathbf{70.17}^{\pm 8.90}$	$\mathbf{62.51}^{\pm 8.44}$	$\mathbf{82.83}^{\pm 7.07}$

The table demonstrates the significant performance improvement achieved by incorporating post-training with OCT images from the Kermany database v3.

Models	Test Acc. ↑	Sensitivity ↑	Specificity ↑
COVID-ViT [8]	$61.40^{\pm 6.40}$	$36.63^{\pm 7.76}$	$61.66^{\pm 8.13}$
JointViT w/o post-training	$57.89^{\pm 0.00}$	$37.45^{\pm 7.76}$	$65.85^{\pm 4.40}$
JointViT w/ post-training	$\mathbf{70.17}^{\pm 8.90}$	$\mathbf{62.51}^{\pm 8.44}$	$\mathbf{82.83}^{\pm 7.07}$

The table showcases the efficacy of balancing augmentation in enhancing model performance, with notable improvements observed in various metrics.

Models	Test Acc. ↑	Sensitivity ↑	Specificity ↑
COVID-ViT [8]	$61.40^{\pm 6.40}$	$36.63^{\pm 7.76}$	$61.66^{\pm 8.13}$
JointViT w/o bal-aug	$64.91^{\pm 2.48}$	$45.63^{\pm 17.55}$	$71.08^{\pm 11.96}$
JointViT w/ bal-aug	$\mathbf{70.17}^{\pm 8.90}$	$\mathbf{62.51}^{\pm 8.44}$	$\mathbf{82.83}^{\pm 7.07}$

CONCLUSION

Oxygen saturation level (SaO2) significantly impacts health, particularly indicating conditions like sleep-related hypoxemia, hypoventilation, sleep apnea, and other related disorders. However, continuous monitoring of SaO2 is time consuming and subject to variation due to patients' status instability. On the other hand, Optical coherence tomography angiography (OCTA) images excel in speed and effectively screening eye-related lesions, showing promise in assisting with the diagnosis of sleep-related disorders. Therefore, we propose a novel method called JointViT, which incorporates a joint loss utilizing both SaO2 values and categories for supervision. Additionally, we employ a balancing augmentation technique during data preprocessing to handle the long-tail nature of OCTA datasets. Our method significantly outperformed other established medical imaging recognition methods, paving the way for the future utilization of OCTA in diagnosing sleep-related disorders.