A Extended classification results

Table 2. Extended classification performance with confidence intervals (CIs) computed at 95% using bootstrapping (n=1000) following [3, 8].

Dataset	ResNet-50			sparse BagNet		
	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
Kaggle	$.935_{(.927-.943)}$	$.765_{(.745-.784)}$	$.967_{(.962-.972)}$	$.904_{(.894-.913)}$	$.706_{(.684-.726)}$	$.978_{(.974-.982)}$
IDRiD	$.864_{(.822-.902)}$	$.963_{(.942-.981)}$	$.714_{(0.639-0.781)}$	$.878_{(.839-.913)}$	$.951_{(.927-.972)}$	$.768_{(.699-.828)}$
E-Ophtha	$.972_{(.958-.986)}$	$.960_{(.929-.987)}$	$.912_{(.877-.944)}$	$.944_{(.921-.966)}$	$.920_{(.882-.957)}$	$.892_{(.852-.928)}$
FGA-DR	$.816_{(.769-.858)}$	$.764_{(.743-.783)}$	$.743_{(.653-.819)}$	$.789_{(.752-.823)}$	$.807_{(.790-.826)}$	$.594_{(.500-.687)}$
DDR	$.963_{(.960-.966)}$	$.801_{(.790-.810)}$	$.973_{(.968-.977)}$	$.926_{(.921-.931)}$	$.668_{(.656-.680)}$	$.980_{(.977-.984)}$
DR2	$.866_{(.825-.905)}$	$.669_{(.591-.742)}$	$.977_{(.959-.993)}$	$.922_{(.889-.951)}$	$.662_{(.584-.742)}$	$.983_{(.968-.997)}$
APTOS	$.972_{(.965-.978)}$	$.942_{(.931-.952)}$	$.956_{(.946-.965)}$	$.995_{(.992-.997)}$	$.982_{(.976-.988)}$	$.965_{(.956-.973)}$
FCM-UNA	$.967_{(.955-.979)}$	$.840_{(.811-.868)}$	$.989_{(.971-1.0)}$	$.935_{(.918-.952)}$	$.698_{(.661-.734)}$	$.989_{(.972-1.0)}$
Messidor-1	$.954_{(.942-.965)}$	$.853_{(.825-.879)}$	$.941_{(.921-.961)}$	$.943_{(.929-.955)}$	$.833_{(.806-.860)}$	$.960_{(.941-.976)}$
Messidor-2	$.925_{(.912-.938)}$	$.794_{(.763-.823)}$	$.893_{(.875-.913)}$	$.876_{(.859-.895)}$	$.751_{(.720-.786)}$	$.887_{(.868-.908)}$
Nakuru Kenya	$.715_{(.692-.739)}$	$.357_{(.324-.394)}$	$.977_{(.974-.980)}$	$.710_{(.687-.733)}$	$.297_{(.265-.331)}$	$.979_{(.976-.982)}$

B Further visualizations

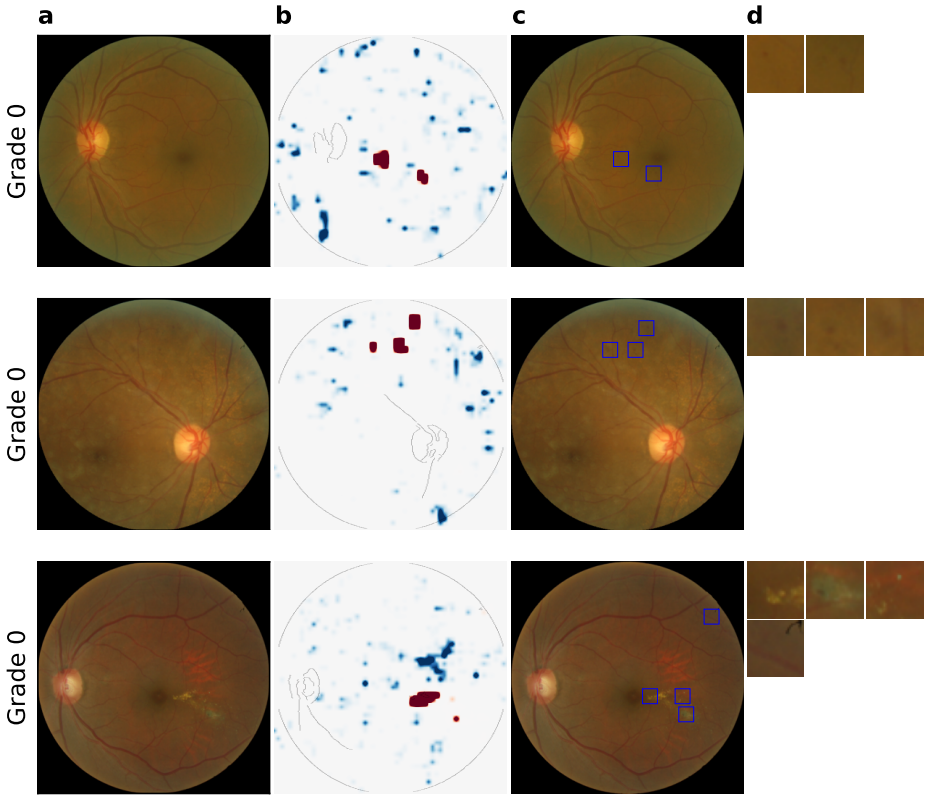

Fig. 3. Examples of false positive prediction with explanation maps.

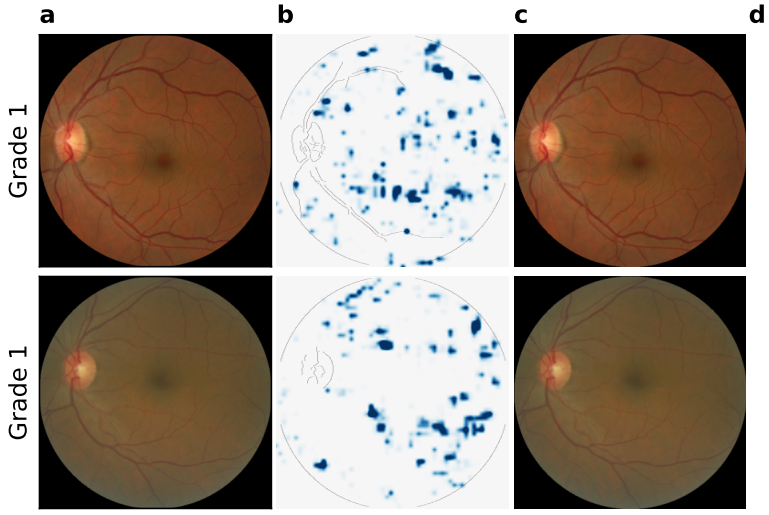

Fig. 4. Examples of false negative prediction with explanation maps.

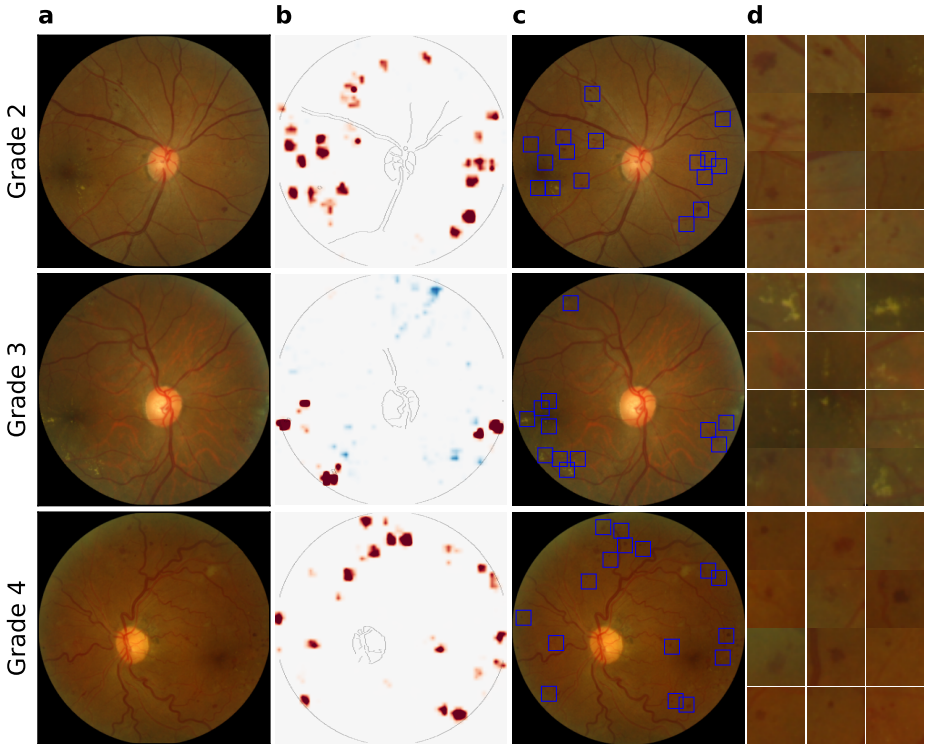

Fig. 5. Examples of True positive prediction with explanations.