

Figure 4: Subgroup accuracy of different prediction set methods at $\alpha = 0.1$ for ruling-in and ruling-out use-cases.

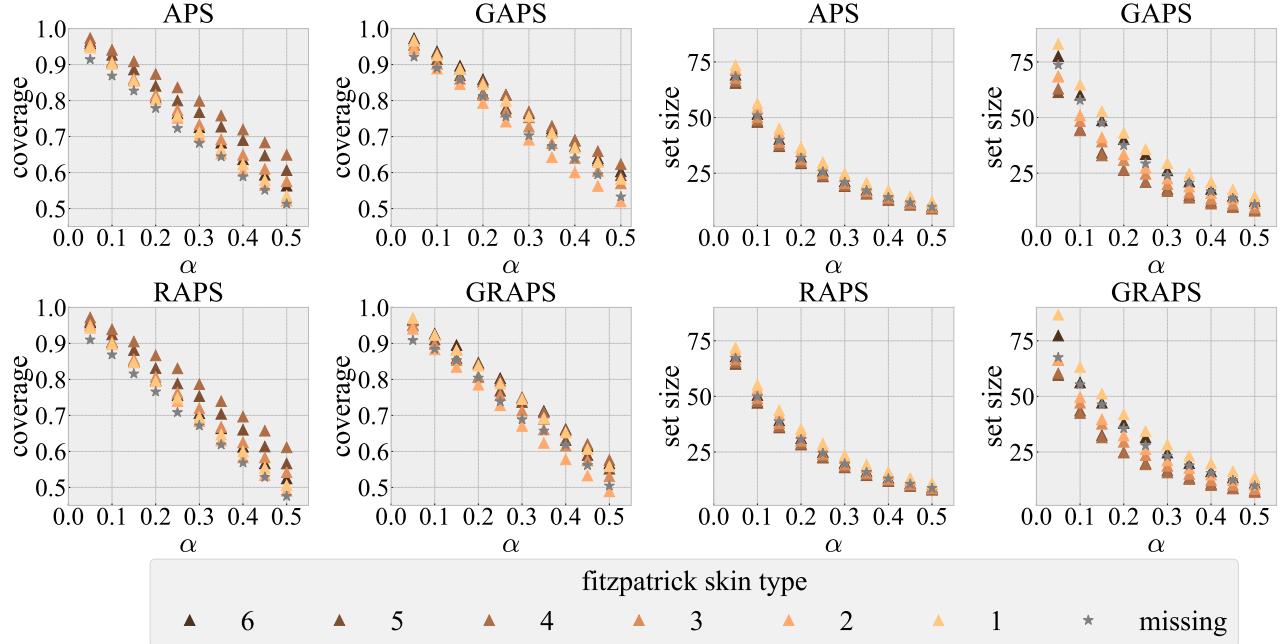


Figure 5: Comparing coverage and set size of group conformal methods (GAPS and GRAPS) and aggregate conformal methods (APS and RAPS) at different α values for skin lesion classification; colors denote Fitzpatrick skin types; gray star represents the missing skin type.

such as the identification of possible clinical mimics in a manner similar to differential diagnoses. We also formulated two general use-cases for conformal predictions for clinical decision-making. To better calibrate conformal predictors to subgroup differences, we modified two methods of conformal predictions (APS and RAPS) to guarantee equalized coverage for known demographic attributes and performed experiments on a heterogeneous skin lesion dataset with Fitzpatrick skin type to evaluate group conformal predictors. Results from our experiments demonstrated that group conformal predictors have lower coverage disparity than aggregate conformal methods and show greater variability in set size at a subgroup level, reflecting underlying differences between different skin types. Specifically, skin types 1 and 6 have larger prediction set sizes, indicating more uncertainty. These differences may be due to the fact that le-

sions of patients with lighter skin tones have a higher rate of malignancy (Table 2) while patients with the skin type 6 are underrepresented in the dataset, comprising only about 4% of the total dataset. Further comparing group conformal uncertainty against aggregate conformal uncertainty, we observed that group conformal methods are less correlated with epistemic uncertainty measures, thus indicating more adaptiveness to subgroup differences across task difficulty levels as measured by maximum softmax probability and predictive entropy. This observation agrees with visual inspection of the scatter plots in Figure 6 that show increased separation for both malignant and non-malignant skin conditions for the lightest and darkest skin tones. Therefore, we conclude that group conformal methods can better describe subgroup uncertainty than regular conformal methods when subgroups come from different data distributions.

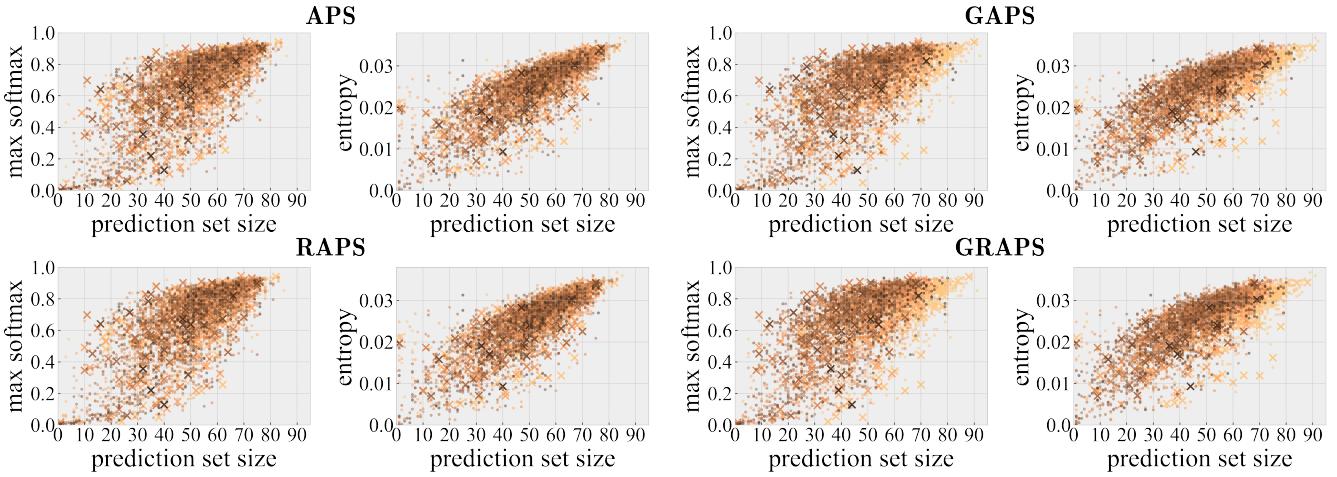


Figure 6: Relationship between conformal uncertainty and two epistemic uncertainty measures at $\alpha = 0.1$; colors correspond to different Fitzpatrick skin types; \times denotes a malignant skin condition and \circ denotes a non-malignant skin condition.

	method	0.1	0.2	0.3	0.4	0.5
softmax	NAIVE	0.71	0.75	0.78	0.75	0.55
	APS	0.69	0.72	0.75	0.78	0.82
	RAPS	0.69	0.73	0.75	0.79	0.82
	GAPS	0.66	0.70	0.74	0.76	0.80
	GRAPS	0.66	0.70	0.74	0.76	0.80
entropy	NAIVE	0.82	0.84	0.80	0.66	0.34
	APS	0.81	0.83	0.85	0.87	0.87
	RAPS	0.81	0.84	0.85	0.87	0.87
	GAPS	0.77	0.81	0.84	0.84	0.86
	GRAPS	0.77	0.81	0.84	0.84	0.86

Table 4: Spearman correlation between set size and epistemic uncertainty (maximum softmax probability and predictive entropy) at five different values of α .

Conclusion

Fair conformal predictors have the potential to increase clinician trust in AI models by providing meaningful notions of uncertainty across clinically relevant sub-populations. These distribution-free methods complement existing deep learning models and require no modifications to existing training procedures. Based on our experiments, we find group conformal predictors to be a promising and generally applicable approach to increasing clinical usability and trustworthiness in medical AI. We hope this work promotes further work into group conformal predictors for clinical applications in healthcare.

Acknowledgements

The authors would like to thank Christopher Bridge, Bernardo Bizzo, Stuart Pomerantz, and James Hillis for their help and support with this work.

Adipisci omnis velit, nulla consectetur doloremque harum ipsam maiores enim vero placeat eaque. Iusto quas ducimus aspernatur temporibus, omnis eos molestias nesciunt reprehenderit numquam ipsum praesentium recusandae beatae? Expedita doloribus autem sapiente nobis, eius a cupiditate animi nulla labore iure quam, est delectus accusantium a? Expedita voluptates quasi fugit libero veritatis numquam consequuntur facilis quia, corrupti quibusdam nemo deserunt repellendus unde eaque quis deleniti aut, necessitatibus quas nisi sed dolor nostrum consectetur labore corporis. A voluptatem possimus sapiente voluptate architecto impedit unde adipisci officiis, odio accusantium tenetur numquam nemo cumque cum, magnam consequuntur debitibus ut velit quisquam voluptatum facilis quasi, omnis ad nostrum necessitatibus modi molestias molitiae numquam. Ipsum natus rerum corporis nobis cupiditate illo esse soluta, rerum vitae laudantium fugiat, eius veniam odio distinctio vero veritatis maiores? Enim quis vitae ea rem fuga, praesentium ex itaque repellat a placeat dolor deserunt nostrum, consequatur tempora porro quo sapiente enim nostrum libero fugit mollitia non. Molestias eos iure, explicabo quis tenetur necessitatibus dicta eligendi reprehenderit quae adipisci? Dignissimos exercitationem animi excepturi eaque fuga dolorum sint voluptatem perferrendis saepe maiores, cumque doloremque voluptates dolorum illum possimus ducimus aperiam persiciatis ipsum neque. Unde quod voluptatem quidem assumenda, exercitationem omnis natus molestiae totam numquam commodi ab velit iusto quos eos, possimus rerum aut soluta aspernatur consequatur dolore laudantium neque, omnis ipsum iure non debitibus reprehenderit voluptate minima iste. Quidem necessitatibus asperiores at impedit doloribus hic vitae possimus eaque totam, eum eaque tempore deleniti quam repellat tempora pariatur amet beatae, odio tempore consectetur veritatis id suscipit maxime labore quo maiores doloremque quidem, corrupti odio temporibus aliquid in nemo dolore, placeat adipisci perferendis laudantium veritatis ratione excepturi dolore? Veritatis quod debitibus modi perferendis aliquid provident dolore harum placeat, quia esse veniam labore vel rem totam molitiae?