(a) from left to right: ViT [5] predictions on image DSC08676 (class "Moderately Dark"), image DSC08549 (class "Medium"), and image DSC07573 (class "Light").

(b) from left to right: ResNet50 [17] predictions on image DSC08676 (class "Moderately Dark"), image DSC08549 (class "Medium"), and image DSC07573 (class "Light").

Fig. 8. Prediction Comparison of ViT and ResNet50.

V. Conclusion

This study has successfully demonstrated the potential of the Vision Transformer model for automating the categorization of coffee bean roast levels. By utilizing computer vision techniques, the proposed system has proven to be both accurate and reliable, achieving a testing accuracy of 0.9778, precision of 0.9791, recall of 0.9778, and F1-score of 0.9777. These results highlight the effectiveness of ViT in distinguishing subtle visual differences between eight distinct roast levels. The system could be used to simplify and accelerate the traditionally manual process of roast level classification, reducing both time and error.

Future work may explore the use of pretrained models for further performance improvements, as well as optimizing ViT for tasks in resource-constrained environments. Additionally, investigating the robustness of these models on larger and more diverse datasets could provide further insights into their applicability for real-world scenarios. For those interested in replicating or extending this study, the code and resources are publicly available [2].

References

[1] G. R. Quadra, E. M. Brovini, J. A. dos Santos, and J. R. Paranaíba, *Caffeine Consumption over Time*. Cham: Springer International Publishing, 2022, pp. 1535–1552. [Online]. Available: https://doi.org/10.1007/978-3-030-92392-1_78

[2] M. Münchow, J. Alstrup, I. Steen, and D. Giacalone, "Roasting conditions and coffee flavor: A multi-study empirical investigation," *Beverages*, vol. 6, no. 2, 2020. [Online]. Available: https://www.mdpi.com/2306-5710/6/2/29

[3] J. L. Massey, *Coffee: Production, Consumption and Health Benefits*. Nova Publishers, 2016.

[4] J. Baggenstoss, L. Poisson, R. Kaegi, R. Perren, and F. Escher, "Coffee roasting and aroma formation: application of different time- temperature conditions," *Journal of agricultural and food chemistry*, vol. 56, no. 14, pp. 5836–5846, 2008. [Online]. Available: https://doi.org/10.1021/jf800327j

[5] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. [Online]. Available: https://openreview.net/forum?id=YicbFdNTTy

[6] J. Sarino, M. Bayas, E. Arboleda, E. Guevarra, and R. Dellosa, "Classification of coffee bean degree of roast using image processing and neural network," *Int. J. Sci. Technol. Res*, vol. 8, no. 10, pp. 3231–3233, 2019.

[7] D. H. Suryana and W. K. Raharja, "Applying artificial intelligence to classify the maturity level of coffee beans during roasting," *International Journal of Engineering, Science and Information Technology*, vol. 3, no. 2, pp. 97–105, 2023. [Online]. Available: https://doi.org/10.52088/ijesty.v3i2.461

[8] X. Gao, Y. Qian, and A. Gao, "Covid-vit: Classification of covid-19 from ct chest images based on vision transformer models," 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2107.01682

[9] DiFluid, "Difluid omni introduction," 2023, [Online]. Available: https://baristagroup.com.au/wp-content/uploads/2023/11/DiFluid_Omni_Introduction.pdf.

[10] ——, "Difluid omni user manual," 2023, [Online]. Available: https://digitizefluid.com/en-au/blogs/coffee/unveiling-the-roast-degree-analysis-technology-of-difluid-omni-1.

[11] D. Shen, "Image registration by local histogram matching," *Pattern Recognition*, vol. 40, no. 4, pp. 1161–1172, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320306003608

[12] P. Sane and R. Agrawal, "Pixel normalization from numeric data as input to neural networks: For machine learning and image processing," in *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2017, pp. 2221–2225. [Online]. Available: https://doi.org/10.1109/WiSPNET.2017.8300154

[13] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. [Online]. Available: https://doi.org/10.1109/5.726791

[14] V. R. Joseph, "Optimal ratio for data splitting," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 15, no. 4, pp. 531–538, 2022. [Online]. Available: https://doi.org/10.1002/sam.11583

[15] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2019. [Online]. Available: https://doi.org/10.48550/arXiv.1711.05101

[16] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: an overview," 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2008.05756

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. [Online]. Available: https://doi.org/10.1109/CVPR.2016.90

[2] GitHub