Tabel 5. Accuracy and F-1 Score for balanced dataset

Method	Accuracy	F-1 Score Macro Avg
KNN	93.67%	93.55%
Random Forest	97.76%	97.70%
Boosted KNN	97.94%	98.00%

4. Conclusion

This research has shown that using imbalanced and balanced datasets gives different results. Based on Table 4, when the methods classify the imbalanced dataset, all methods have good accuracy. However, the F-1 score macro average of every method indicates that the performance of the models needs to be improved. The imbalanced data training resulted in the models classifying most cases as the majority class, so the models do not perform well on the rare class or the minority class. Table 5 shows that the accuracy of the balanced dataset is lower than that of the imbalanced dataset. However, each method's F-1 score macro average is significantly improved rather than the imbalanced dataset. Therefore, the performance of every method for both classes has improved. Apart from good results of F-1 scores, Table 5 shows a comparison between KNN, RF, and BK. The KNN method has 93.67% of accuracy and 93.55% of F-1 score. Meanwhile, the Random Forest has 97.76% of accuracy and 97.70% of F-1 score. Thus, this result indicates that Random Forest performs better than the KNN method. On the other hand, the BK model (combining the concept of Random Forest and Ensemble Learning) achieves 97.94% of accuracy and 98.00% of F-1 score. Therefore, the BK has succeeded in boosting the performance of the original KNN method. Future work in developing a machine learning model for classification can be related to implementing the RF concept to another machine learning method, such as Naive Bayes or Logistic Regression. Besides that, the BK model can be implemented into text or image classification.

Daftar Pustaka

[1] A. Almomany, W. R. Ayyad, and A. Jarrah. Optimized implementation of an improved knn classification algorithm using intel fpga platform: Covid-19 case study. *Journal of King Saud University-Computer and Information Sciences*, 34(6):3815–3827, 2022.

[2] D. D. Astuti, R. B. Adriani, and T. W. Handayani. Pemberdayaan masyarakat dalam rangka stop generasi stunting. *JMM (Jurnal Masyarakat Mandiri)*, 4(2):156–162, 2020.

[3] M. H. Barri, F. Alia, L. Novamizanti, R. Purnamasari, F. Akhyar, T. Fahrudin, P. H. Gunawan, and S. Mandala. Aksi cegah stunting melalui aplikasi sagita: Status gizi balita. *JMM (Jurnal Masyarakat Mandiri)*, 7(2), 2023.

[4] R. G. Devi and P. Sumanjani. Improved classification techniques by combining knn and random forest with naive bayesian classifier. In *2015 IEEE international conference on engineering and technology (ICETECH)*, pages 1–4. IEEE, 2015.

[5] R. Devika, S. V. Avilala, and V. Subramaniyaswamy. Comparative study of classifier for chronic kidney disease prediction using naive bayes, knn and random forest. In *2019 3rd International conference on computing methodologies and communication (ICCMC)*, pages 679–684. IEEE, 2019.

[6] N. Dogru and A. Subasi. Traffic accident detection using random forest classifier. *2018 15th Learning and Technology Conference, L and T 2018*, pages 40–45, 2018.

[7] V. Herliansyah, R. Latuconsina, and A. Dinimaharawati. Prediksi stunting pada balita dengan menggunakan algoritma klasifikasi naïve-bayes. *eProceedings of Engineering*, 8(5), 2021.

[8] T. R. Hoens and N. V. Chawla. *Imbalanced Datasets: From Sampling to Classifiers*, chapter 3, pages 43–59. John Wiley & Sons, Ltd, 2013.

[9] V. Jackins, S. Vimal, M. Kaliappan, and M. Y. Lee. Ai-based smart prediction of clinical disease using random forest classifier and naive bayes. *The Journal of Supercomputing*, 77:5198–5219, 2021.

[10] V. Matzavela and E. Alepis. Decision tree learning through a predictive model for student academic performance in intelligent m-learning environments. *Computers and Education: Artificial Intelligence*, 2:100035, 2021.

[11] Ramli, K. E. Agho, K. J. Inder, S. J. Bowe, J. Jacobs, and M. J. Dibley. Prevalence and risk factors for stunting and severe stunting among under-fives in north maluku province of indonesia. *BMC pediatrics*, 9:1–10, 2009.

[12] S. Sagita and K. N. Siregar. Faktor-faktor risiko stunting pada balita di indonesia: Suatu scoping review. *Media Publikasi Promosi Kesehatan Indonesia (MPPKI)*, 5(6):654–661, 2022.

[13] A. S. M. Sohail and P. Bhattacharya. Classification of facial expressions using k-nearest neighbor classifier. In *Computer Vision/Computer Graphics Collaboration Techniques: Third International Conference, MIRAGE 2007, Rocquencourt, France, March 28-30, 2007. Proceedings 3*, pages 555–566. Springer, 2007.

[14] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip. A comparison of random forest variable selection methods for classification prediction modeling. *Expert systems with applications*, 134:93–101, 2019.

[15] H. H. Sutarno, R. Latuconsina, and A. Dinimaharawati. Prediksi stunting pada balita dengan menggunakan algoritma klasifikasi k-nearest neighbors. *eProceedings of Engineering*, 8(5), 2021.

[16] M. A. U. H. Tahir, S. Asghar, A. Manzoor, and M. A. Noor. A classification model for class imbalance dataset using genetic programming. *IEEE Access*, 7:71013–71037, 2019.

[17] C. R. Titaley, I. Ariawan, D. Hapsari, A. Muasyaroh, and M. J. Dibley. Determinants of the stunting of children under two years old in indonesia: A multilevel analysis of the 2013 indonesia basic health survey. *Nutrients*, 11(5):1106, 2019.

Lampiran