Enhancing Tree Classification Performance with Convolutional Neural Networks Using Weighting, Augmentation, and Intricate Model Techniques

Soichiro Tanabe, 2 May 2024

1. Background and context

It has been reported that forests in certain areas, such as Germany, have faced severe challenges due to climate change, which has led to extensive damage and poses long-term threats to forest health and development (Holzwarth et al., 2020). Urgent action is required to address these issues, including implementing adaptive forest management practices and investing in research and innovation for forests, as emphasised by Holzwarth et al. (2020). Therefore, sufficient forest data is crucial for effective forest monitoring. The Global Forest Resources Assessment (FRA) 2015 succeeded in improving forest resource data from 112 countries, covering about 83% of the global forest area, indicating a positive trend towards reduced rates of forest loss and carbon emissions, along with an increase in the capacity for sustainable forest management (MacDicken et al., 2015).

In addition to utilising forest data for monitoring, machine learning and Artificial Intelligent offer potential solutions for this endeavour; however, the challenge lies in acquiring, preparing, and maintaining a high-quality dataset for effective deep learning model training (Hamedianfar et al., 2022). Ahlswede et al. (2023), despite challenges within the dataset, demonstrated the potential of deep learning techniques to support forestry administration in generating large-scale tree species maps to address challenges driven by global environmental change.

2. Aim and Objectives

In this research, similar to Ahlswede et al. (2013), the same dataset, the TreeSatAI Benchmark Archive, was adopted. This dataset is designed to gather multi-sensor and multi-label information for the classification of 20 tree species in central Europe, including 50,381 aerial images of publicly managed forests located in Lower Saxony, Germany (Ahlswede et al., 2022). The study aimed to investigate how Convolutional Neural Network (CNN) models perform in predicting and identifying classes of forests for monitoring purposes. Particularly, the study focused on how changing models and modifying datasets impact performance and the differences in predicting classes of trees. To address these questions, two levels were chosen: Level 1 and Level 2 (Figure 1). Additionally, the study examined the differences in performance between the two levels.

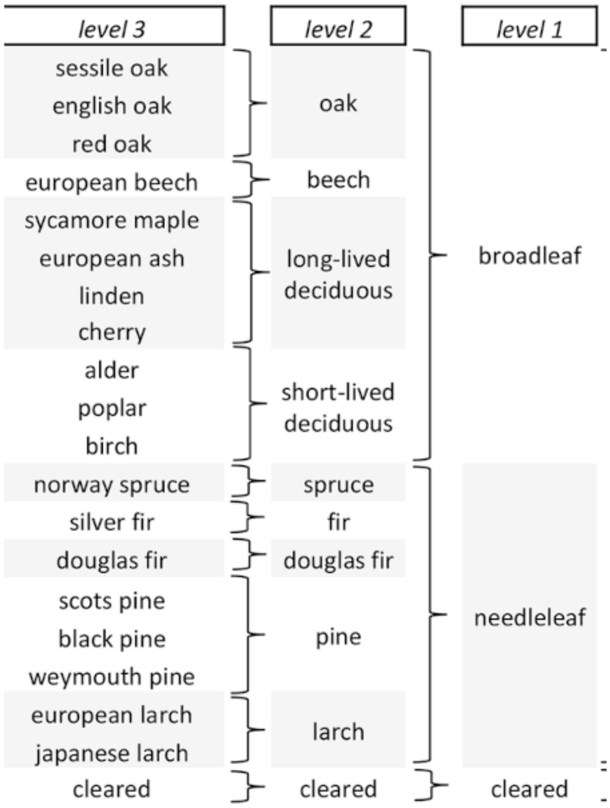


Figure1: Image detail from Ahlswede et al. (2023)

3. Methods

3.1 Level 1 investigation

The tree classes in Level 1 are divided into three groups: broadleaf, needleleaf, and cleared. The dataset was divided into training, validation, and test data sets to apply CNN models. Notably, Figure 2, which illustrates the distribution of the three unique classes in the training data, demonstrates that there is an extremely smaller number of data points for 'cleared' compared to the other two classes of data.

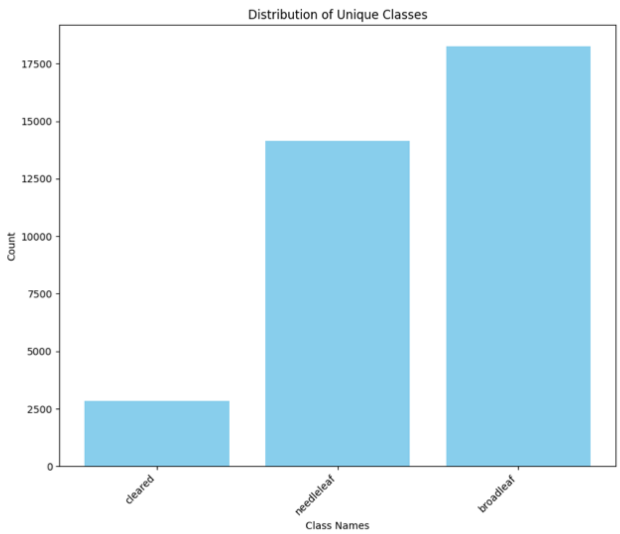


Figure 2

3.1.1 Base model

As the initial model, a simple CNN model was set up, with the following structure:

```
Conv2D(8, (5, 5), padding='same', activation='relu', input_shape=input_shape),
Conv2D(8, (5, 5), padding='same', activation='relu'),
MaxPooling2D(pool_size=(4, 4)),
Flatten(),
Dense(128, activation='relu'),
Dense(num_classes, activation='softmax')
```

Based on the initial research settings, it was discovered that the disparity in accuracy and loss between the training and validation datasets indicates overfitting. The further detailed discovery of this model will be mentioned in the section of 'result'.

3.1.2 Implement a more intricated model

Implementing a more intricate model involves adding additional layers to enable the neural network to capture more intricate patterns within the data, potentially enhancing its predictive capabilities (Saturn Cloud, 2023). Ramesh(2018) states that increasing the number of filters in each layer enhances the depth of the feature space, enabling the CNN to learn more levels of global abstract structures. Additionally, incorporating dropout is beneficial for mitigating overfitting, as demonstrated by the base model, as it encourages the network to learn more robust and generalized representations of the data (The Open University, 2023a). Therefore, a more intricate model was structured as follows:

```
Conv2D(16, (5, 5), padding='same', activation='relu', input_shape=input_shape),
Conv2D(16, (5, 5), padding='same', activation='relu'),
MaxPooling2D(pool_size=(4, 4)),
Conv2D(32, (3, 3), padding='same', activation='relu'),
Conv2D(32, (3, 3), padding='same', activation='relu'),
MaxPooling2D(pool_size=(2, 2)),
Conv2D(64, (3, 3), padding='same', activation='relu'),
Conv2D(64, (3, 3), padding='same', activation='relu'),
MaxPooling2D(pool_size=(2, 2)),
Flatten(),
Dense(128, activation='relu'),
Dropout(rate=0.5),
Dense(num_classes, activation='softmax')
```

3.1.3 Add weighting

Figure 2 depicts imbalanced class data, which can profoundly affect classification algorithms, resulting in biased performance (The Open University, 2023a). To address this issue, techniques such as over-sampling (e.g., SMOTE) or weighting data can be used; however, the class weight approach is considered to be more efficient than over-sampling techniques (The Open University, 2023b). Therefore, a technique involving weighting imbalanced data was applied to modify the training data. Subsequently, the previous two models utilised this modified training data for training.

3.2 Level 2 investigation

The trees in Level 2 are divided into more detailed groups, comprising ten distinct categories. Similar to the setup in Level 1, the train data, validation data, and test data were structured accordingly. Additionally, the number of epochs and the optimizer settings were configured in the same manner as Level 1.

3.2.1 Base model

As the base model, the simple model was set as the same as the structure as described in 3.1.1.

3.2.2 Implement a more intricated model

The more intricate model was structured similarly to that in 3.1.2, aiming to enable the neural network to learn more levels of global abstract structures and capture more intricate patterns within the data, as indicated by Ramesh. S (2018) and Saturn Cloud (2023) as well as to encourage the network to learn more robust and generalised representations of the data, as indicated by the Open University (2023a).

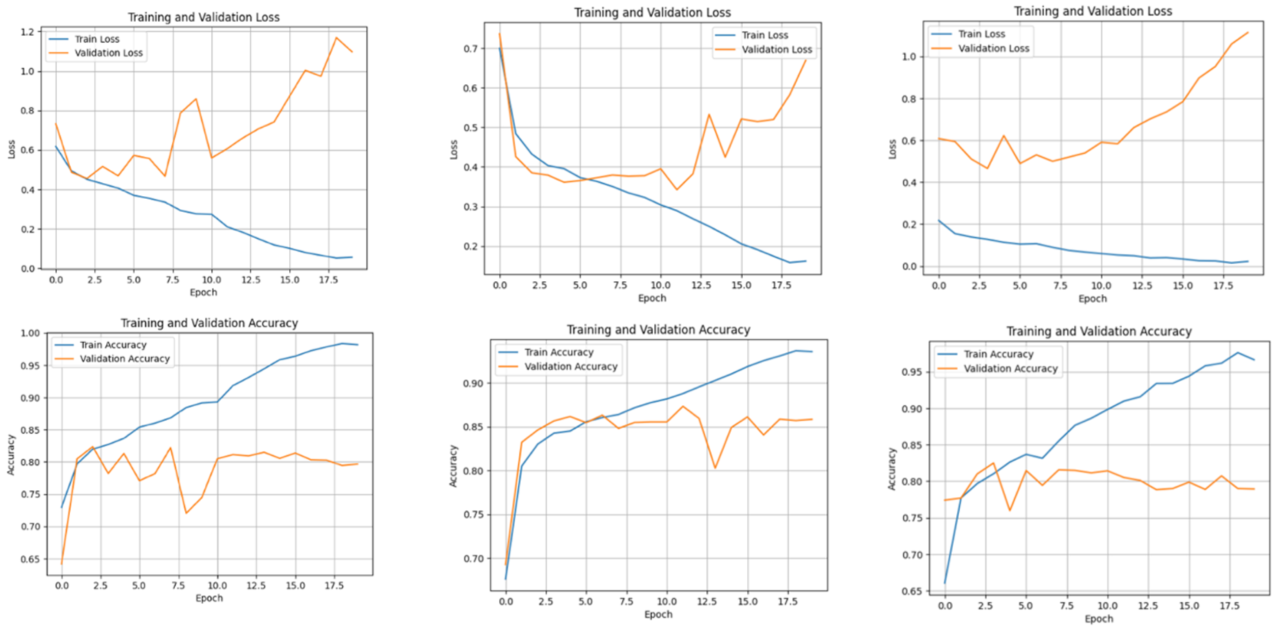
3.2.3 Data Augmentation

Although the detailed results will be mentioned in the 'Results' section, it is worth noting that overfitting was observed in both the base and intricate models (3.2.3). To mitigate overfitting, while changing the dropout rate is one solution, another useful technique is data augmentation, which offers the generation of more training data in the case of a lack of data (Open University, 2023a). This time, since dropout has already been added and with the aim of obtaining more data variety, modified training data using random rotation up to 10% and random zoom up to 20% in data augmentation were applied to the intricate model (3.2.2).

4. Result

4.1 Level 1

4.1.1 : the loss and accuracy for both the training and validation datasets



3.1.1 base model

3.1.2 intricate model

3.1.3 Add weighting

Figure 3

Figure 3 illustrates the loss and accuracy for both the training and validation datasets in the base model, the intricate model, and the model with added weighting. All three models displayed a dissociation between the training and validation data in terms of loss and accuracy, indicating overfitting. Generally, the lines for both training and validation closely aligned up to around epochs 2 and 3, but diverged thereafter. However, in the model with added weighting, there was a significant gap between the training and validation loss lines, which never aligned.

4.1.2 Overall performance

	Accuracy	*Loss*
Base Model	78.87%	1.1165
Intricate Model	85.16%	0.6259
Model with Added Weights	84.76%	0.4415

Table 1

Table 1 displays accuracy and loss metrics for each model. The Intricate Model achieved the highest accuracy (85.16%), followed by the Model with Added Weights (84.76%) and the Base Model (78.87%). In terms of loss, the Model with Added Weights had the lowest (0.4415), followed by the Intricate Model (0.6259) and the Base Model (1.1165). Overall, the Intricate Model and the Model with Added Weights outperformed the Base Model, exhibiting higher accuracy and lower loss.

4.1.3 Precision and recall

	Base Model		Intricate Model		Model with Added Weights	
	precision (%)	recall (%)	precision (%)	recall (%)	precision (%)	recall (%)
cleared	8.99	9.13	7.83	6.43	**9.55**	**10.37**
needleleaf	40.45	35.63	40.7	37.49	39.93	38.17
broadleaf	**52.5**	**57.12**	52.22	56.94	51.9	52.88

Table 2

Across all models, it was evident that predicting the 'cleared' class was challenging, with precision and recall values hovering around 8 to 10%. Conversely, predicting the 'broadleaf' class was relatively easier, exceeding 50%. Specifically, the Model with Added Weights demonstrated superior performance in both precision and recall for the 'cleared' class (9.55% and 10.37%, respectively), while the base model exhibited the best performance for the 'broadleaf' class (52.5% and 57.12%, respectively).

4.1.4 True Positives

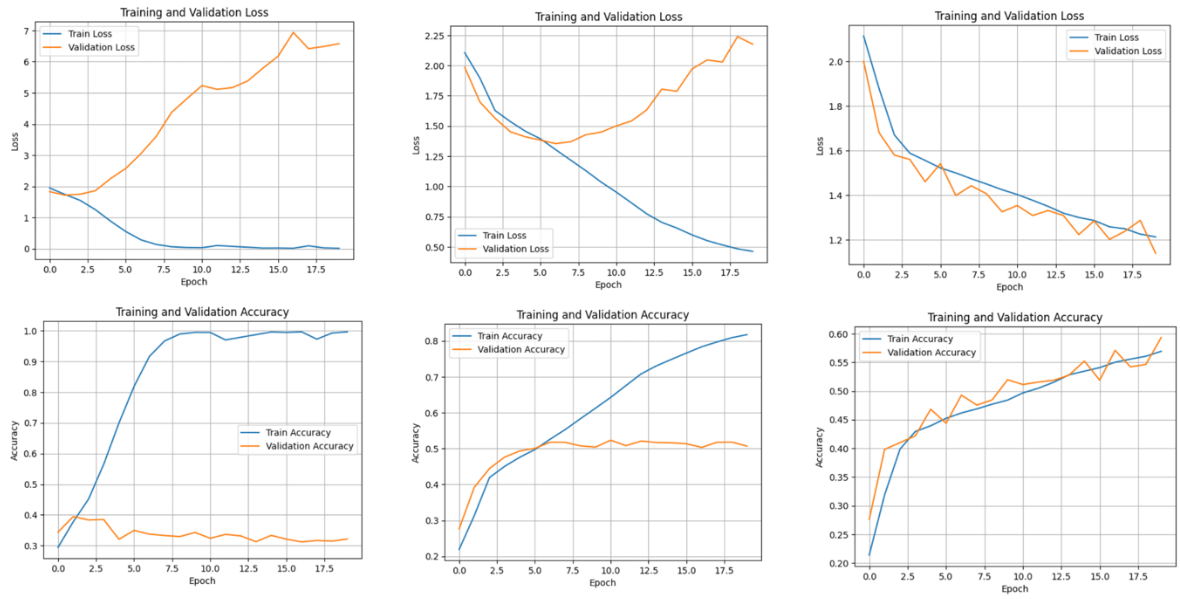
	cleared	***needleleaf***	***broadleaf***
Base Model	81	1414	**2982**
Intricate Model	57	1488	2973
Model with Added Weights	**92**	**1515**	2761

Table 3

Table 3 shows the number of True Positives for each class and each model. Through the technique of adding weights, the number of 'cleared' classs, belonging to the minority class, increased by 12.35% compared to the Base Model. Similarly, the number of 'needleleaf' classs also increased by 7.14%. However, concerning 'broadleaf', the Base Model achieved the highest number of True Positives, reaching 2982.

4.2 Level 2

4.2.1 : the loss and accuracy for both the training and validation datasets



3.2.1 base model 3.2.2 intricate model 3.2.3 data augmentation

Figure 4

Figure 4 reveals overfitting in both the Base and Intricate Models, evidenced by the disparity between training and validation data in Loss and Accuracy. While the Base Model shows a significant gap in both metrics, the Intricate Model initially aligns but later diverges after 5 epochs. In contrast, the Model with Augmentation displays better alignment, despite some fluctuations in validation lines for both Loss and Accuracy.

4.2.2 Overall performance

	Accuracy	*Loss*
Base Model	31.58%	6.7639
Intricate Model	50.77%	2.1504
Model with Augmentation	59.84%	1.1308

Table 4

Table 4 summarises the overall performance of each model. The Base Model had the lowest accuracy (31.58%) and highest loss (6.7639), indicating its struggle in classification. Conversely, the Intricate Model showed improvement with higher accuracy (50.77%) and lower loss (2.1504). The Model with Augmentation performed best, boasting the highest accuracy (59.84%) and lowest loss (1.1308), likely due to its augmentation techniques enhancing generalisation.

4.1.3 Precision and recall

	Base Model		Intricate Model		Model with Augmentation	
	precision (%)	recall (%)	precision (%)	recall (%)	precision (%)	recall (%)
larch	7.24	4.24	7.07	7.38	7.27	1.88
cleared	10.2	8.00	8.58	9.47	7.73	7.67

beech	12.34	14.22	13.24	14.61	11.64	13.51
oak	15.20	15.38	15.04	13.66	**16.11**	**23.17**
douglas fir	5.77	5.72	6.07	7.38	5.78	4.24
pine	13.13	14.29	14.51	17.51	14.17	16.76
fir	1.05	0.53	1.49	0.53	**0**	**0**
short-lived deciduous	10.02	10.09	12.2	13.72	11.14	7.96
long-lived deciduous	12.33	16.95	10.64	6.00	10.56	7.43
spruce	11.22	8.65	10.31	10.31	12.66	15.91

Table 5

Across all models, it was noticeable that the percentages of precision and recall were very low, not achieving more than 30%. In the Model with Augmentation, the class 'oak' achieved the highest percentages among the three models (precision: 16.11%, recall: 23.17%), while the lowest percentages of precision and recall, both 0%, were observed in the class 'fir'. Overall, the analysis of precision and recall data suggests that the performance of each model varies across different classes, with no single model consistently outperforming the others across all classes.

5. Evaluation

5.1 Level 1

To mitigate unbalanced classes, the technique of adding weights was effective in boosting predictions for the minority class, 'cleared'. Additionally, the class 'needleleaf' had the highest number of True Positives. On the other hand, the class 'broadleaf' was not identified as frequently as in the Base Model. This suggests that the model focused on identifying the sparsely distributed classes rather than the majority class, resulting in a slight trade-off. Although techniques like increasing filters, adding layers, dropout, and adding weights improved accuracy and loss significantly compared to the base model, overfitting remained a challenge.

5.2 Level 2

The Model with Augmentation achieved the best performance among models, significantly increasing by 20.26% compared to the base model, while also mitigating overfitting. The Level 2 research demonstrated that the techniques of increasing filters, adding layers, dropout, and augmentation effectively worked. However, the accuracy is moderately low, not reaching more than 60% and indicating room for further improvement. Additionally, the Model with Augmentation observed 0% precision and recall for the class 'fir', despite achieving the highest percentages for the class 'oak'.

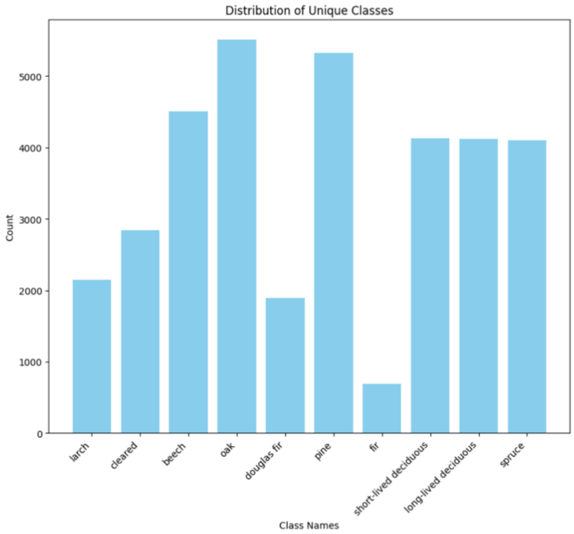


Figure 5: The distribution of the three unique classes in Level 2

Examining Figure 5, the distribution of classes for 'oak' is the highest, while the distribution for 'fir' is the lowest. This indicates that the Model with Augmentation prioritised predicting classes with higher distribution, potentially at the expense of lower ones. Consequently, accuracy percentages increased, even though 'fir' predictions were absent. Despite the inability to predict 'fir' at all, the accuracy percentages remained higher.

5.3 Comparison between Level 1 and Level 2

The Base Model and Intricate Model were applied to both Level 1 and Level 2. It is evident that the models performed significantly better on Level 1 compared to Level 2. Specifically, the accuracy in Level 1 surpassed 80%, whereas in Level 2, it remained below 55%. This discrepancy suggests that predicting classes at Level 1, with fewer classes, yields superior results compared to Level 2. Therefore, for identifying classes in the TreeSatAI Benchmark Archive, it is evident that working with a smaller number of classes leads to more accurate predictions.

6. Discussion of wider implications

Through the research conducted at Level 1 and Level 2, it became evident that the Intricate Model is effective in improving performance. The technique of adding weights to mitigate unbalanced classes in Level 1 not only showed effectiveness in identifying minority classes but also helped mitigate overfitting. However, it is not sufficient to conclude that the performance with this technique is always better than without it. Regarding augmentation, the research conducted at Level 2 showed improvements in performance. Despite these improvements, the overall performance remains suboptimal, highlighting the necessity for further refinement through hyperparameter tuning and other strategies. Additionally, continuous testing and monitoring in production environments are recommended to mitigate harm during model deployment.

Comparison with the research by Ahlswede et al. (2023) revealed the use of additional models such as ResNet and LightGBM, which demonstrated higher performances. Therefore, further investigation with different models, as well as adjustments to hyperparameters, needs to be considered for further improvement.

7. Conclustion

This research evaluated model performance in predicting tree classes across Level 1 and Level 2 datasets. It was discovered that adding weights effectively mitigated class imbalance at Level 1, while augmentation techniques showed improved performance at Level 2 as well as reducing overfitting despite focusing on predicting classes that are more common, potentially neglecting the less prevalent ones. Furthermore, intricate models incorporating techniques such as adding layers, dropout, and increasing filters consistently demonstrated improved performance across both Level 1 and Level 2 datasets.

While Level 1 exhibited superior performance due to its fewer classes, indicating the importance of dataset characteristics in model design, there is still room for improvement. Future research needs to focus on integrating additional models and fine-tuning hyperparameters to enhance predictive accuracy and reliability in tree classification tasks, particularly in handling imbalanced class distributions. This will assist in monitoring the wide array of tree types, including those with imbalanced distributions in real-world scenarios, in order to tackle challenges arising from global environmental change.

(words: 2452)

8. Reference

Ahlswede, S., Schulz, C., Gava, C., Helber, P., Bischke, B., Förster, M., Arias, F., Hees, J., Demir, B. and Kleinschmit, B. (2022) ' TreeSatAI Benchmark Archive: a multi-sensor, multi-label dataset for tree species classification in remote sensing', *Earth System Science Data*, 8 February 2022, pp.1-22 [Online]. Available at https://essd.copernicus.org/articles/15/681/2023/ (Accessed 11 May 2024).

Hamedianfar, A., Mohamedou, C., Kangas, A., and Vauhkonen, J.(2022) 'Deep learning for forest inventory and planning: a critical review on the remote sensing approaches so far and prospects for further applications', *Forestry,* 95, 451–465 [Online]. Available at https://www.researchgate.net/publication/358220604_Deep_learning_for_forest_inventory_and_planning_A_critical_review_on_the_remote_sensing_approaches_so_far_and_prospects_for_further_applications (Accessed 11 May 2024).

Holzwarth, S., Thonfeld, F., Abdullahi, S., Asam, S., Da Ponte Canova, E., Gessner, U., Huth, J., Kraus, T., Leutner, B., Kuenzer, C. (2020) 'Earth Observation Based Monitoring of Forests in Germany: A Review', *MDPI*, 31 October 2020 [Online]. Available at https://doi.org/10.3390/rs12213570 (Accessed 11 May 2024).

MacDicken, K., Jonsson, Ö., Piña, L., Maulo, S., Contessa, V., Adikari, Y., Garzuglia, M., Lindquist, E., Reams, G., and D'Annunzio, R. (2015) 'Global Forest Resources Assessment 2015. How are the World's Forests Changing? (Second edition)', *Food and Agriculture Organization of the United Nations*, pp. 7-12 [Online]. Available at https://openknowledge.fao.org/server/api/core/bitstreams/29b8ae23-99f9-4a05-b796-9a35d02af29d/content (Accessed 11 May 2024).

Ramesh. S (2018) ' A guide to an efficient way to build neural network architectures- Part II: Hyper-parameter selection and tuning for Convolutional Neural Networks using Hyperas on Fashion-MNIST', *Medium*, 7 May 2018 [Online]. Available at https://towardsdatascience.com/a-guide-to-an-efficient-way-to-build-neural-network-architectures-part-ii-hyper-parameter-42efca01e5d7 (Accessed 11 May 2024).

Saturn Cloud (2023) *How to Improve Accuracy in Neural Networks with Keras*, 6 July 2023 [Online] Available at https://saturncloud.io/blog/how-to-improve-accuracy-in-neural-networks-with-keras/ (Accessed 11 May 2024).

Schulz, C., Ahlswede, S., Gava, C., Helber, P., Bischke, B., Arias, F., Förster, M., Hees, J., Demir, B., and Kleinschmit, B. (2022) ' TreeSatAI Benchmark Archive for Deep Learning in Forest Applications', *Zenodo*, 1 August 2022 [Online]. Available at https://zenodo.org/records/6780578 (Accessed 11 May 2024).

The Open University (2023a) 5 Training of CNNs, *TM358 Weeks 8-11 Block 2: Image recognition with CNNs* [Online]. Available at https://learn2.open.ac.uk/mod/oucontent/view.php?id=2152484&section=5.1 (Accessed 11 May 2024).

The Open University (2023b) 5 Imbalanced data, *TM358 Weeks 24 Block 6: Data handling* [Online]. Available at https://learn2.open.ac.uk/mod/oucontent/view.php?id=2217059&section=5.5.2.2 (Accessed 11 May 2024).