Knowledge distillation from multispectral Images for fish freshness estimation

G. Tsagkatakis*, *S. Roumpakis**, *S. Nikolidakis*†, *E. Petra*‡, *A. Mantes*‡, *A. Kapantagakis*†, *K. Grigorakis*†, *G. Katselis**, *N. Vlahos**, *P. Tsakalides**

* *Institute of Computer Science, Foundataion for Reseach and Technology - Hellas (FORTH), 70013, Heraklion, Crete, Greece*
† *Hellenic Centre for Marine Research, 19013, Anavyssos, Attiki, Greece*
‡ *Information Management Systems Institute, Athena Research Center, 15125, Marousi, Greece*
* *Department of Fisheries and Aquiculture Technology, University of Patras, 30200, Messolonghi, Greece*

Abstract

Fish quality is primarily effected by the number of days elapsed since harvesting, while bad storage conditions can also lead to quality degradation similar to the impact time. Existing approaches require laboratory testing, a laborious and time-consuming process. In this work, we investigate technologies for quantifying fish quality though the development of deep learning models for analyzing imagery of fish. We first demonstrate that such a quantification is possible, to a certain degree, from multispectral images provided a sufficient number of training examples is available. Given that, we explore how knowledge distillation can be utilized for achieving similar fish quality estimation accuracy, but instead of using high-end multispectral imaging systems, using off-the-shelf RGB cameras. Experimental evaluation on individuals from the Mullus Marbatus family demonstrates that the proposed methodology constitutes a valid approach.

Introduction

European legislation categorizes fish into four categories based on their quality value as food, namely Excellent (class A), Category A (class B), Category B (class C), and Inadmissible (class C). In the first category, the samples are classified almost immediately after their catch while in the latter category, they are unsuitable for human consumption [1]. The classification of each sample to each class is based on the Posthumous enzyme activity is responsible for the degradation in fish quality due to the breakdown of cell membranes and is influenced by several factors such as the type of fish, the way of extinction, and the method of preservation.

In this work, we consider the FRESQO platform, a compact and portable system which will consist of an compact multispectral imaging camera and an embedded processing system, while the interfacing with the user is achieved through two-way communications to his/her smartphone. The proposed platform offers sever benefits compared to more traditional approaches such as the extremely fast evaluation, the ability to perform in-situ estimation, and the non-destructive nature of the method. A detailed description of the system characteristics and design specifications is analysed in [10], while an illustration of the platform is shown in Figure 1.

While the merits of the proposed platform have been previously discussed, the platform assumes the availability of Hyperspectral Imaging (HSI) observation over 25 spectral bands in the visible to near-infrared range. This requirement imposes hard
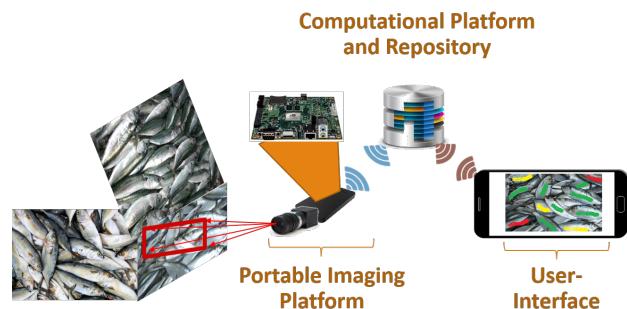

Figure 1. Overview of the FRESQO platform.

constraints on the different imaging options, introducing significant costs in developing such as a platform. The aim of this work is to demonstrate that training data collected from such platforms can be utilized by deep learning models to offer significant performance improvement when training deep learning models with a significantly smaller number of spectral bands. We achieve this objective through a process known as Knowledge Distillation (KD) [2] which involves transferring part of the feature extraction from the full spectral resolution network (teach network) to the limited resolution network (student) as shown in Figure 2.
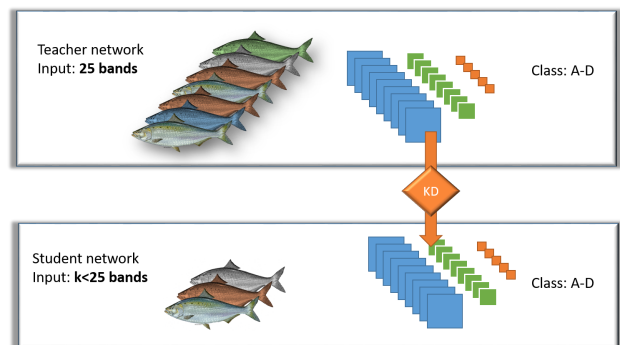

Figure 2. Knowledge distillation process where a teacher network is trained with full spectral resolution observation and a student using both low spectral resolution examples and features extracted from the teacher network.

State-of-the-art

Degradation in fish quality is due to the posthumous enzyme activity which leads to the breakdown of cell membranes which in turn leads to the exponential increase in the bacterial load, causing the severe degradation of the skin. Currently, different types of instruments are available for the quantification of fish freshness which exploits the alteration of different physiological parameters. Traditional approaches for estimating the freshness of a fish either employs human senses like appearance, odor, flavor, and texture or rely on laboratory tests. These techniques are either subject o a very high degree of subjectivity or require specialized personnel and are often destructive. During the past decades, optical imaging technologies have been readily considered for the evaluation and inspection of food quality, including multispectral imaging and near-infrared spectroscopy. Different approaches that have been presented include methods based on enzyme biosensors, electrochemical biosensors, colorimetric sensors, electronic tongue, and different types of spectroscopy[3]. Multispectral imaging, a particular case of spectroscopy is among the most prominent solutions for this problem due to the numerous benefits such methods offer including in-situ and real-time estimation. In this work, we employ Snapshot Spectral Imaging technologies which in addition to the general benefits of spectroscopy, do not require scanning the item over a conveyor belt. This technology has been recently employed for classifying generic objects [7], as well for food quality monitoring including detecting plant diseases [8] and red-meat classification [9].

Method

The objective of this paper is the demonstration of the capabilities of state-of-the-art deep learning architectures and more specifically Convolutional Neural Networks (CNN) in analysing multispectral images for estimating the freshness of fish. By fish freshness, we mean the estimation of the apparent days since harvest, which may not coincide with the actual number of days in cases of problematic storage conditions. The targeted scenario involves a compact and portable system which will consist of a snapshot spectral imaging system and a mini-pc system like the NVIDIA Jetson, while the interfacing with the user will be achieved through transmission of the processed images to his/her smartphone, as shown in Figure 1.

The proposed system currently consists of a snapshot spectral imaging camera and a data processing pipeline based on machine learning for estimating the elapsed time between harvesting and imaging. Specifically, observations are acquired over the visible-near IR range (400-1000nm) using a snapshot spectral camera from Photon Focus (MV1-D2048x1088-HS02-96-G2), equipped with an IMEC sensor acquiring snapshot spectral imagery of 2048×1088 pixels at 42fps. Unlike traditional linescan approaches, can acquire the entire spectral profile of a scene from a single image (exposure). This capability makes snapshot spectral cameras ideal for deployment in real-time environments, removing the need for specialized platforms like conveyor platforms. The quantum efficiency of each band is shown in Figure 3.

We employ a two-step process where we first train a network, the teacher, using multispectral images as input and then train another network, the student, using RGB imagery. Specifically, the teacher network is a Convolutional Neural Network (CNN) that
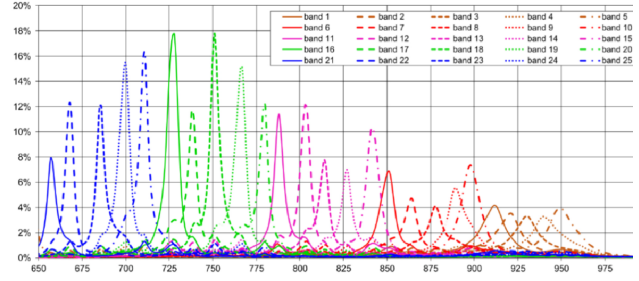

Figure 3. *Quantum efficiency associated with each spectral band.*

accepts as input a 25 spectral band cube and predicts one out of five classes of fish freshness as the classification output. Once, the teacher network is trained, we train another CNN which accepted 3 color band cubes as inputs and similarly predicted the fish freshness class. To explore the knowledge acquired by the teacher network, the student network trained such that (i) makes accurate predictions and (ii) the predicted logits are close to the ones predicted by the teacher network.

Formally, we define a single input example as a three dimensional cube $\mathbf{X} \in \mathbb{R}^{w \times h \times k}$, where w, h, k are the width, height and number of bands. The target variable, corresponding to the fish quality metric, is given by $\mathbf{y}$ which is a n-dimensional vector encoding the corresponding label in a one-hot encoding scheme. We define a deep learning network as the function $\mathscr{F}$, which takes an input spatio-spectral cube and produces a fish freshness index

$$\hat{\mathbf{y}} = \mathscr{F}(\mathbf{x}; \mathbf{w}) \tag{1}$$

where $\mathbf{w}$ are the weights on the network.

The objective in training such a network is to minimize an appropriately defined loss function, such as the categorical cross-entropy given by:

$$\min_{\mathbf{w}} - \sum_{i=1}^{N} \mathbf{y} \log \hat{\mathbf{y}} \tag{2}$$

where N is the number of training examples.

The goal of this work is to exploit features, extracted from the neural network when trained using the full spectral resolution observations from k bands, in order to increase the capabilities of similar networks when trained on a subsection of the spectral resolution $k' < k$. To that end, define the original network $\mathscr{F}$ as the teacher network while the resolution spectral resolution network $\mathscr{G}$ as the student network. Furthermore, we define the features extracted from input $\mathbf{x}$ at a particular layer i as $\mathscr{F}_i(\mathbf{x})$ for the teacher and $\mathscr{G}_i(\mathbf{x}')$ where $\mathbf{x}'$ corresponds to the spectral subsampled version of $\mathbf{x}$.

To train the student network, we employ both the categorical cross entropy, as in the teacher, as well as a feature similarity terms, penalizing the difference between features extracted from the teacher and the student network given the corresponding inputs. Specifically, the student network loss function is given by:

$$\min_{\mathbf{w}} - \sum_{i=1}^{N} \mathbf{y} \log \hat{\mathbf{y}} + \tau \| \mathscr{F}_i(\mathbf{x}) - \mathscr{G}_i(\mathbf{x}') \|_2 \tag{3}$$

where τ controls the importance of the knowledge distillation term.

For the particular problem we consider in this work, the acquired spatio-spectral image cube is introduced into a Convolutional Neural Network (CNN) which consists of four Convolutional layers and one Global Max Pooling layer. The specifics of the considered architecture are shown in Figure 4. A significant difference between the proposed architecture and typical architectures employed in image classification is that the spatial size of the image is rectangular and that the multispectral images consist of 25 spectral bands, much higher compared to 3 in color imagery.

Layer (type)	Output Shape	Param #
Input Layer	[(None, 216, 409, 23, 1)]	0
Conv3D	(None, 216, 409, 1, 8)	192
Batch Normalization	(None, 216, 409, 1, 8)	32
Reshape	(None, 216, 409, 8, 1)	0
MaxPooling3D	(None, 7, 15, 8, 1)	0
Flatten	(None, 840)	0
Dropout	(None, 840)	0
Dense	(None, 4)	3364

Total params: 3,588
Trainable params: 3,572
Non-trainable params: 16

Figure 4. *Network architecture including dimensions for each layer and associated number of parameters.*

Experimental results

We evaluate the method on fish from the Mullus Marbatus family over four days since harvesting, thus generating the 4 class. We performed two sets of daily acquisition of multispectral images from 5 individuals. To train and validate the performance of the system, individuals from the first set were considered as training examples and from the second set as validation examples. Both the training and the validation set images (spectral cubes) were augmented through geometric transform (translation, rotation, and clipping) leading to 128 training examples and 32 validation.

In the experimental analysis, we consider three distinct scenarios, namely (i) Train/validate Teacher Network with 25 spectral bands (Hyperspectral), (ii) Train/validate of Student with 12 spectral bands (Multispectral), and (iii) Train/validate Student network with 3 spectral bands (RGB). While for training the Teacher Network, a single training/validation approach is available, for the student networks there are two strategies. Specifically, the two strategies are:

- Naïve, where the student is training using only the currently available data (5 bands in MSI and 3 in RGB).
- Knowledge Distillation (KD), where training utilizes both using available data, as well as the similarity to features between student and teacher network.

Training Teacher Network on Hyperspectral data

The first step in applying the proposed KD approach involves training the teacher network using the full resolution observations. The overall prediction accuracy is 98.4% for the training
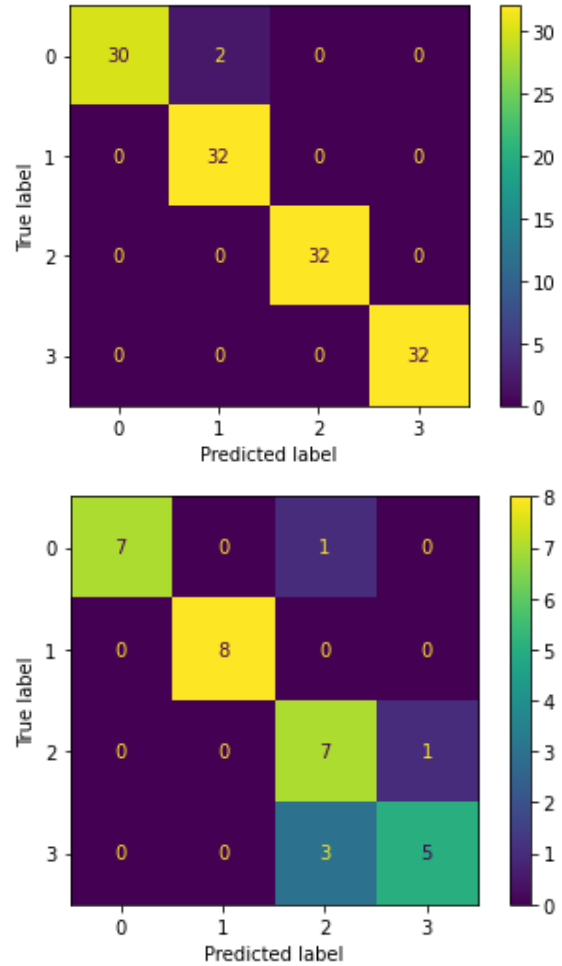
and 84.3% for the validation.

Figure 5. *Confusion matrix for training (top) and validation (bottom) set using 23 spectral bands as the input.*

Performance is also presented in terms of confusion matrices in Figure 5 where one can observe that the majority of misclassifications corresponds to predicting a adjacent class, e.g. instead of class A, predict class B.

Multispectral

Given a fully training teacher network, we argue that learned features can be utilized for training the student network. The performance on the validation for the naive and KD students is presented in Figure 6. These results demonstrate two things, first, that KD can lead to higher accuracy (78.1%) compared to the naive approach (71.8%), while for the case of misclassification, the errors from the KD are more constrained around the diagonal of the matrix, i.e., the errors are typically between adjacent classes.

To further quantify the performance gain of the KD during the training of the student network, Figure 7 present the accuracy as a function of training epoch for the naive and KD approach respectively. Given these results, we can make two observations. First, both training and validation accuracy is higher when em-
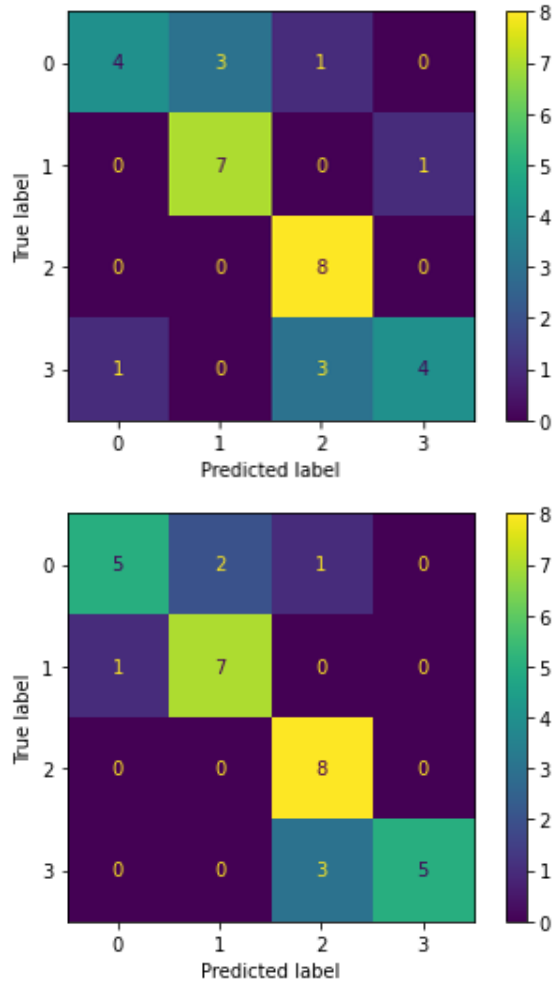
Figure 6. *Confusion matrix on validation set for the naive student (top) and the KD student (bottom) using 12 spectral bands as the input.*

ploying the KD approach compared to the naive method. Second, the performance for the case of KD, although increasing with more training epochs, starts from a significantly higher performance point compared to the naive case. This demonstrates that employing the KD can significantly increase performance when limitations in terms of training resources are present.

RGB

Similar to the previous case, we also explored the cased of KD for training networks using 3 spectral bands as a representative example of an RGB-based system. The confusion matrices for the naive and KD student training are presented in Figure 8 while Figure 9 presents the classification accuracy as a function of training epoch. Comparing the case of naive and KD student training, there is a significant increase in terms of classification accuracy, from 65.6% to 84.3% respectively when employing the KD approach. For the case of classification accuracy, we can, similarity to the case of MSI, observe that employing KD can significantly boost the accuracy achieved when only three color channels are available.
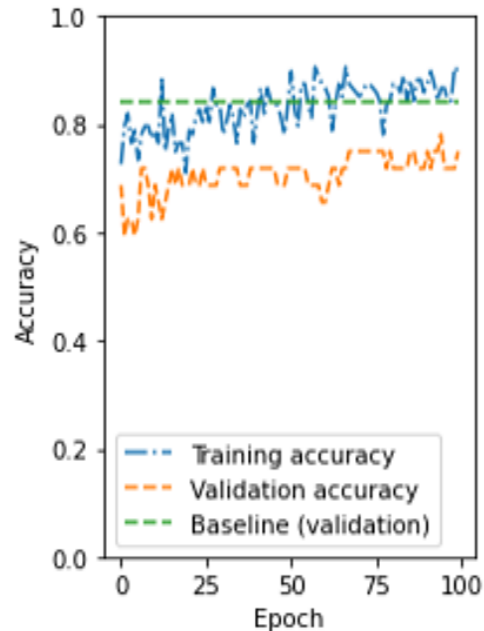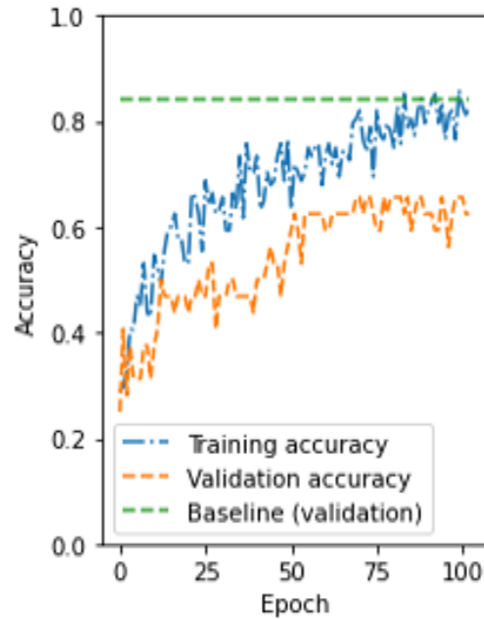

Figure 7. *Classification accuracy for the naive student (top) and the KD student (bottom) using 12 spectral bands as the input.*

Conclusions

In this paper, we present our initial results on the utilization of features extracted from hyperspectral imagery for the estimation of fish freshness from multispectral and typical color images. To that end, we consider the introduction of knowledge distillation with a deep learning algorithm for feature transferring. The results indicate that by introducing information for hyperspectral data, the system can achieve very high prediction accuracy even using as little as 3 color channels.
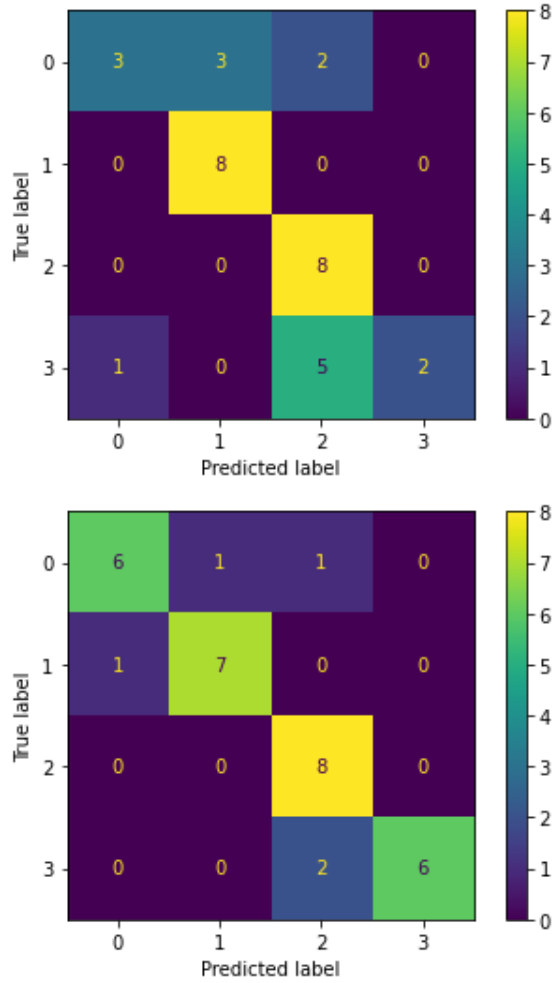
Figure 8. *Confusion matrix on validation set for the naive student (top) and the KD student (bottom) using 3 spectral bands as the input.*

Acknowledgments

References

[1] Cheng, J.H., Sun, D.W., Han, Z. and Zeng, X.A., 2014. Texture and structure measurements and analyses for evaluation of fish and fillet freshness quality: a review. Comprehensive Reviews in Food Science and Food Safety, 13(1), pp.52-61.

[2] Hinton, Geoffrey, Oriol Vinyals, and Jeff Dean. "Distilling the knowledge in a neural network." arXiv preprint arXiv:1503.02531 (2015).

[3] Wu, L., Pu, H. and Sun, D.W., 2019. Novel techniques for evaluating freshness quality attributes of fish: A review of recent developments. Trends in food science & technology, 83, pp.259-273.

[4] LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. nature, 521(7553), pp.436-444.

[5] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In Advances in
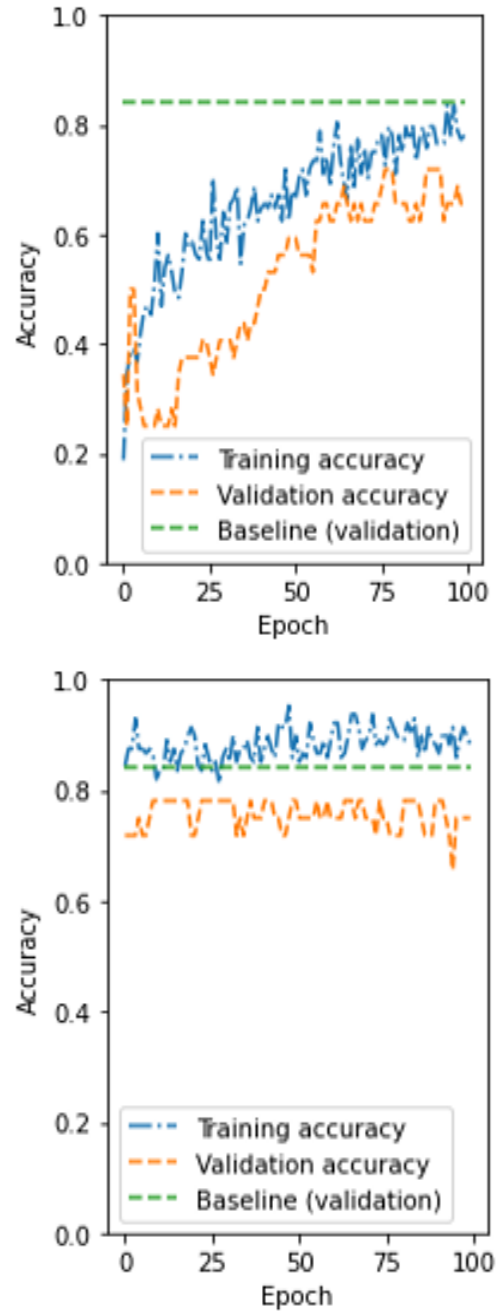

Figure 9. *Classification accuracy for the naive student (top) and the KD student (bottom) using 3 spectral bands as the input.*

neural information processing systems (pp. 1097-1105).

[6] Martinsdóttir, E., Sveinsdottir, K., Luten, J., Schelvis-Smit, R. and Hyldig, G., 2001. Sensory evaluation of fish freshness. IJmuiden, The Netherlands: QIM eurofish.

[7] Fotiadou, K., Tsagkatakis, G. and Tsakalides, P., 2017. Deep convolutional neural networks for the classification of snapshot mosaic hyperspectral imagery. Electronic Imaging, 2017(17), pp.185-190.

[8] Nagasubramanian, K., Jones, S., Singh, A.K., Singh, A., Ganapathysubramanian, B. and Sarkar, S., 2018. Explaining hyperspectral

imaging based plant disease identification: 3D CNN and saliency maps. arXiv preprint arXiv:1804.08831.

[9] Al-Sarayreh, M., Reis, M.M., Yan, W.Q. and Klette, R., 2018, November. Deep spectral-spatial features of snapshot hyperspectral images for red-meat classification. In 2018 International Conference on Image and Vision Computing New Zealand (IVCNZ) (pp. 1-6). IEEE.

[10] Tsagkatakis, G., S. Nikolidakis, E. Petra, A. Kapantagakis, K. Grigorakis, G. Katselis, N. Vlahos, and P. Tsakalides. "Fish Freshness Estimation though analysis of Multispectral Images with Convolutional Neural Networks." Electronic Imaging 2020, no. 12 (2020): 171-1.

[11] Al-Sarayreh, Mahmoud, Marlon M. Reis, Wei Qi Yan, and Reinhard Klette. "Deep spectral-spatial features of snapshot hyperspectral images for red-meat classification." In 2018 International Conference on Image and Vision Computing New Zealand (IVCNZ), pp. 1-6. IEEE, 2018.