




## Article

# Classification of Histamine Content in Fish Using Near-Infrared Spectroscopy and Machine Learning Techniques

Duy Khanh Ninh <sup>1,\*</sup> , Kha Duy Phan <sup>2</sup>, Cong Tuan Vo <sup>3</sup>, Minh Nhat Dang <sup>3</sup>  and Nhan Le Thanh <sup>2</sup> 

<sup>1</sup> Faculty of Information Technology, The University of Danang—University of Science and Technology, Danang 550000, Vietnam

<sup>2</sup> Danang International Institute of Technology (DNIIT), The University of Danang, Danang 550000, Vietnam; pdkha@dniit.udn.vn (K.D.P.); nhan.le-thanh@univ-cotedazur.fr (N.L.T.)

<sup>3</sup> Faculty of Chemical Engineering, The University of Danang—University of Science and Technology, Danang 550000, Vietnam; vctuan@dut.udn.vn (C.T.V.); dmnhat@dut.udn.vn (M.N.D.)

\* Correspondence: nkduy@dut.udn.vn

**Abstract:** Near-infrared (NIR) spectroscopy has emerged as a popular technique for assessing food quality due to its advantages over complex chemical analysis methods. However, the application of NIR spectroscopy for evaluating fish quality based on histamine content has not been extensively explored. This study investigates the use of NIR spectroscopy in combination with machine learning (ML) techniques to classify fish samples into two safety classes, Safe and Unsafe, based on their histamine content. A comprehensive NIR dataset comprising 11,360 spectra collected at eight distinct positions within the fish body was obtained from 284 fish samples of mackerel, tuna, and pompano species. ML experiments were conducted to classify fish samples based on whether their histamine content exceeded the permissible limit of 100 ppm. To address class imbalance and optimize ML models, various data pre-processing and feature extraction techniques as well as ML algorithms were explored. The results demonstrated that utilizing NIR data specifically obtained from the tail's flesh, a specific location within the fish, yielded superior models for fish safety classification. A feature extraction method employing pre-processed NIR spectra and their second derivatives, combined with an optimized convolutional neural network architecture, outperformed traditional ML classifiers with an accuracy of approximately 93%.

**Keywords:** near-infrared spectroscopy; histamine content; fish quality assessment; nondestructive analysis; machine learning; convolutional neural network



**Citation:** Ninh, D.K.; Phan, K.D.; Vo, C.T.; Dang, M.N.; Le Thanh, N. Classification of Histamine Content in Fish Using Near-Infrared Spectroscopy and Machine Learning Techniques. *Information* **2024**, *15*, 528. <https://doi.org/10.3390/info15090528>

Academic Editor: Gabriele Gianini

Received: 16 July 2024

Revised: 16 August 2024

Accepted: 20 August 2024

Published: 1 September 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Considering that fish quality directly affects consumers' health status, we should put a significant amount of importance on assessing fish quality. Furthermore, since the consumption of fish has been on the rise in many parts of the world, and the hygiene and safety of food is of increasing interest to agencies and customers, solving the problem of fish quality control is becoming even more urgent in this context.

Histamine is an endogenous toxin commonly formed in many types of fish. The formation of histamine is a result of the improper storage of fish at incorrect temperatures and durations, which can cause illness in consumers. Histamine poisoning from seafood is primarily associated with the consumption of tuna, herring, anchovies, sardines, and mackerel. In these fish species, certain bacteria can synthesize the enzyme histidine decarboxylase. This enzyme catalyzes the reaction that converts histidine into histamine. Once histamine is formed, it cannot be eliminated by heat (including cooking) or freezing [1].

The permissible limits for histamine content in fish are regulated according to country and region [2]. Australia and New Zealand allow a maximum histamine level in a fish sample of 100 mg/kg (or ppm). The maximum allowable level in Europe is 100 ppm to 200 ppm, while in the USA it must not exceed 50 ppm. In Vietnam, according to the

standard on tuna raw material named TCVN 12153:2018 [3], the histamine level in tuna must not exceed 100 ppm. In this study, we chose a permissible histamine limit of 100 ppm in a fish sample to align with the regulations of Vietnam and many other countries.

Chemical analysis is an effective tool to determine the presence of histamine in fish. Since histamine is often unevenly distributed within fish or batches of fish, the reliability of histamine analysis depends on the sampling method. A large sample size is required. The method of collecting fish samples is also very important. Regarding histamine analysis, the challenge is to completely separate histamine from a large number of interfering substances like histidine or carnosine. Most methods require elaborate and careful processing to remove potential interferents, thus extending the analysis time [2]. Therefore, there is a need to develop rapid analysis methods.

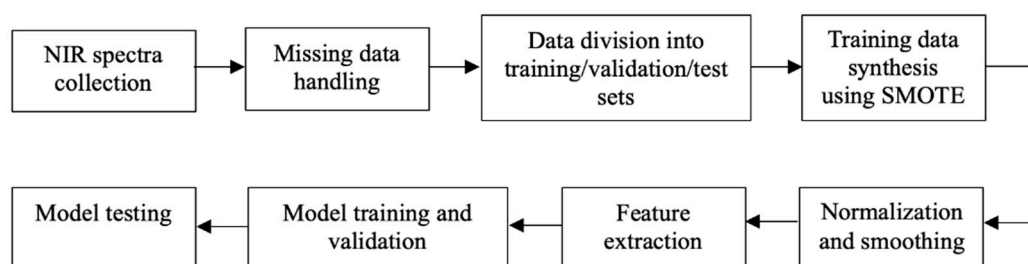
Among the rapid histamine content analysis methods, biosensors are quantitative analytical tools consisting of a biologically based sensor component integrated with a physicochemical transducer. These devices utilize specific biochemical reactions mediated by isolated enzymes, immunosystems, tissues, organelles, cells, and analyze chemical compounds usually through electrical, thermal, or optical signals. Some market products include BIOFISH 300 and BIOFISH 700. Besides biosensors, other rapid analysis methods include colorimetric methods, such as enzyme kits, Hista strip, Agra strip, and ELISA methods using XL665-labeled histamine and Cryptate-labeled antibodies. Biosensors can measure on site, with a relatively quick analysis time compared to other methods. While an ELISA has better sensitivity, colorimetric methods have the shortest analysis time [4].

In recent years, significant advancements in low-cost handheld near-infrared (NIR) spectrometers and machine learning/deep learning (ML/DL) techniques have created new opportunities for the development of rapid, non-destructive, and cost-effective histamine analysis methods. Near-infrared (NIR) spectroscopy coupled with ML has emerged as a promising tool for the non-destructive and rapid assessment of fish quality, both quantitatively and qualitatively [5]. Recent studies have explored its potential in predicting various fish attributes such as freshness [6–10], fat content [11–13], and species identification [14,15]. In addition, NIR spectroscopy and ML techniques have found application in food quality control across various food industries beyond fish. Examples of successful applications include the detection of adulteration in lamb, beef [16], and milk [17], the identification of unauthorized preservation techniques in fermented sausages [18], the discovery of spoilage bacteria in pork [19], the exposure of mislabeling related to production processes in eggs [20], the geographical origin of honey [21], and other aspects of fruits [22–24]. ML algorithms, including partial least squares regression, support vector machines, and artificial neural networks, have been extensively employed for spectra analysis and modeling in past studies. However, challenges persist in terms of data pre-processing, model optimization, and the need for larger, more diverse datasets to enhance the generalization capability of the developed models.

Although official methods for histamine testing in fish and seafood [25] are generally accurate, specific, precise, and well established, they have some drawbacks. These include the high cost of instruments, facilities, and reagents, the need for large amounts of solvents and samples, the destructive nature of some analyses, the requirement for extensive sample preparation and/or post-treatment steps, long analysis times, and the need for skilled operators. Given the need to enhance fish safety controls and transition towards risk-based inspection protocols, the adoption of advanced, rapid, and efficient food inspection technologies could significantly augment well-established methods. Therefore, the present study aims to pioneer the application of NIR spectroscopy and ML techniques for the direct classification of the histamine content of raw fish samples as either safe (below 100 ppm) or unsafe without requiring sample destruction. This approach promises substantial benefits in terms of streamlined workflows and sample conservation for fish industries and markets, where timely food safety information is critical for effective management and loss prevention. Additionally, competent authorities could leverage this technology to bolster their inspection capabilities.

## 2. Materials and Methods

Figure 1 presents the complete workflow of our study. The flowchart begins with the collection of NIR spectra, followed by the handling of missing data to ensure a complete dataset. The data are then divided into training, validation, and test sets, facilitating the construction of a robust predictive model. Notably, the workflow incorporates the SMOTE (Synthetic Minority Over-sampling Technique) [26], which synthesizes additional training data to address class imbalance in the dataset. Subsequently, the data undergo normalization and smoothing processes to ensure consistency and reduce noise. Feature extraction is then performed to identify and prioritize relevant information, enhancing the model's performance. The extracted features are utilized for training and validating the machine learning model, enabling fine-tuning of its parameters. Finally, the model's performance is evaluated using an independent test set, ensuring its generalization ability to unseen data.



**Figure 1.** The complete workflow of our study.

In the following sub-sections, we will describe each step in the workflow in detail.

### 2.1. Data Collection

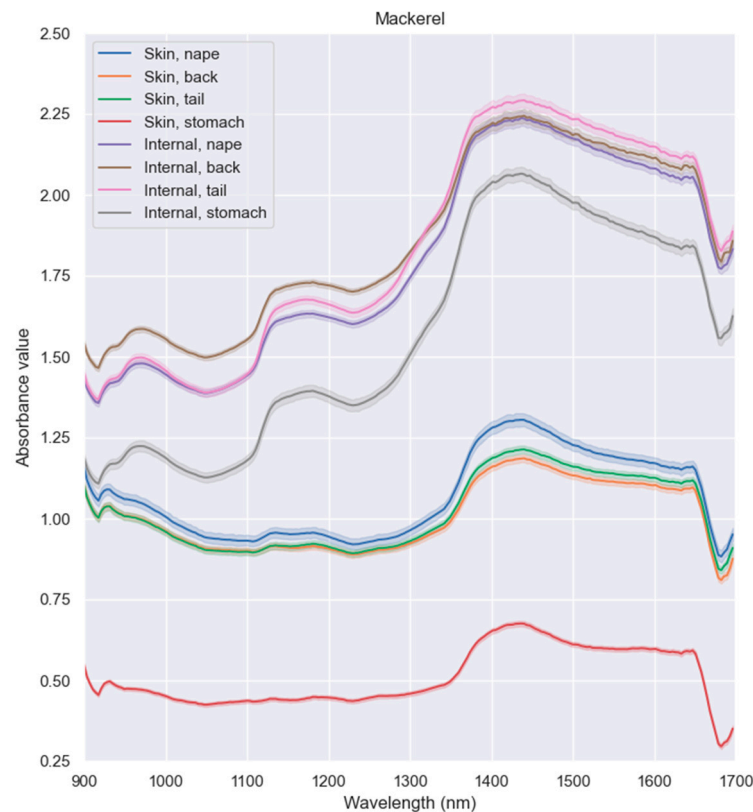
We used a low-cost handheld NIR device, which is the DLP NIRscan Nano EVM produced by Texas Instruments, to measure the NIR spectra of fish samples (Figure 2). In this process, a portion of the radiation in the NIR range emitted by the device was absorbed by the dissected samples. The remaining radiation, which was not absorbed, was either reflected back to the device sensor or transmitted through the substances. According to this, we could achieve absorbance, reflectance, and transmittance spectra simultaneously. Each spectrum consists of 228 wavelengths in the range of 900–1700 nm, i.e., a resolution of 3.5 nm per wavelength point. Among the three types of spectra, we decided to use the absorption spectrum to conduct the experiments for the classification of histamine content in fish.



**Figure 2.** NIR spectrum measurement on a fish's body.

The dataset used in this study includes 11,360 samples of NIR absorption spectra from 284 fish samples (107 mackerel, 109 tuna, and 68 pompano samples). The fish samples, after being collected from the market, were divided into two groups. One group was preserved in cold conditions of 4 °C and samples were taken for analysis at the following time points: 6 h, 12 h, 18 h, and 24 h. The other group was exposed to an atmosphere at

ambient temperature and samples were taken for analysis at the following time points: 0 h, 4 h, 8 h, 12 h, and 16 h. Each fish sample for analysis was subjected to NIR spectrum measurements at four points on the outside skin, including the nape, back, stomach, and tail, as well as four points inside the flesh at the positions of the nape, back, stomach, and tail. Each of the eight positions was measured five times. Consequently, a fish sample produced 40 different NIR spectrum samples. Figure 3 illustrates the representative spectra of the mackerel in terms of the mean and 95% confidence interval of absorbance values aggregated over all mackerel samples in the dataset according to measurement positions on the fish body. It can be seen that the absorbance values vary significantly across different parts of the mackerel. Similar trends were also observed for the tuna and pompano.



**Figure 3.** The absorbance values vary significantly across different parts of the mackerel.

## 2.2. Data Labeling and Division

After collecting NIR data, the fish was filleted, minced, and then its histamine content was determined accurately using a standard analytical chemistry method, which is high-pressure liquid chromatography or HPLC. Forty NIR spectra were then assigned a safety label according to the histamine content of the fish sample. In the event that the histamine content of the fish fell below the established threshold of 100 ppm, the NIR spectra associated with that fish were classified as “Safe”; conversely, if the histamine content exceeded the permissible limit, the NIR spectra were classified as “Unsafe”. Figure 4 illustrates the average NIR absorption spectra of fish samples in the dataset with respect to fish types and safety labels.

Finally, the whole NIR spectrum dataset was divided into three subsets including training, validation, and test sets at the ratio of 3:1:1 for the training, validating, and evaluating classification models. The data division was carried out so as to satisfy the following two criteria: a fish sample and its associated spectrum samples only belonged to one subset, and the histamine content distributions of three subsets were similar (as shown in Figure 5). These requirements were met to assure the objectiveness of the model building and evaluation processes.

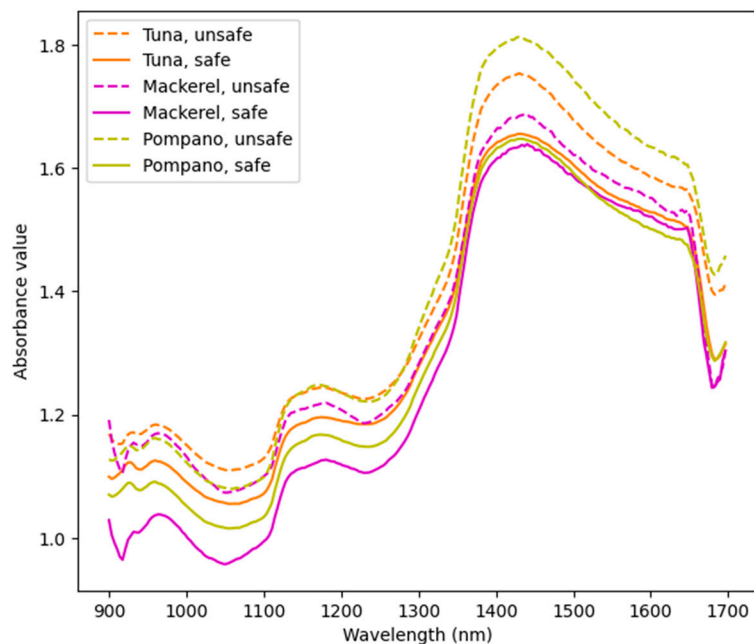


Figure 4. Average NIR absorption spectra of fish samples with respect to fish types and safety labels.

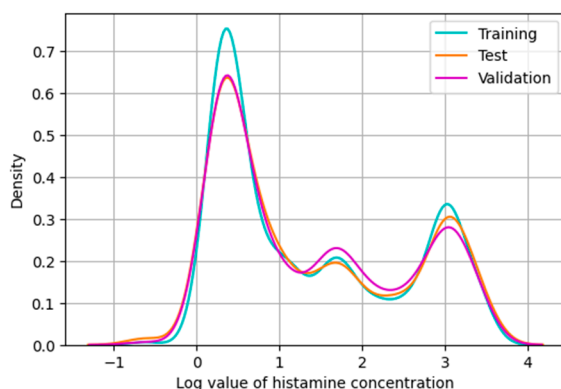


Figure 5. Histamine content distributions of three subsets (X axis uses a base–10 logarithmic scale).

### 2.3. Data Pre-Processing

In the pre-processing stage, we performed three techniques in a row including missing data handling, data normalization, and data smoothing. If a wavelength of an absorption spectrum was missing, the missing absorbance value was replaced by the average of the absorbance values of the two neighboring wavelengths. Then, standard normal variate correction (i.e., z-score normalization) was applied to every single spectrum of the dataset to eliminate the deviations caused by particle size and scattering, making the NIR data consistent. Eventually, the NIR spectra were streamed through a Savitzky–Golay (SG) filter with a window length of 13 points and a polynomial order of 5 to smooth the spectra, thereby removing part of the noise [27]. These parameters of the SG filter were chosen experimentally because they ensured that the resulting spectra were not over-smoothed and that important spectral characteristics remained.

What is especially notable about our dataset is the severe imbalance between the two safety classes. The number of NIR samples belonging to the “Safe” class is nearly six times higher than the “Unsafe” class. This can cause a classification model to be biased towards the majority class with the “Safe” label. To solve this problem, we leveraged the SMOTE technique to handle data imbalance. SMOTE specifically generates new data points for the minority class with the “Unsafe” label. It analyzes existing minority data points and generates new ones similar to them. By adding these synthetic samples, SMOTE balances



the data, giving the model a better capability to learn the minority class. After applying SMOTE on the training subset, the number of NIR samples belonging to the “Unsafe” class is equal to that of the “Safe” class. The details of the SMOTE algorithm can be found in [26]. The synthetic spectrum samples were also normalized and smoothed in the same way as the original ones.

#### 2.4. Feature Extraction

Relevant features need to be chosen for building classification models. For a fish sample, its pre-processed NIR spectrum is a certain choice for the feature vector for safety classification. We further examined the derivatives of the pre-processed spectrum to see if they can help to differentiate labels of safety. We investigated six types of feature vectors based on the concatenation of the pre-processed spectrum and its derivatives as described in Table 1.

**Table 1.** Six feature types of the NIR spectrum.

| Feature Type | Vector Size    | Description                                 |
|--------------|----------------|---|
| orig         | $228 \times 1$ | Original spectrum                           |
| prep         | $228 \times 1$ | Pre-processed spectrum                      |
| der1         | $228 \times 1$ | 1st derivative of pre-processed spectrum    |
| der2         | $228 \times 1$ | 2nd derivative of pre-processed spectrum    |
| prep + der1  | $456 \times 1$ | Pre-processed spectrum + its 1st derivative |
| prep + der2  | $456 \times 1$ | Pre-processed spectrum + its 2nd derivative |

#### 2.5. Model Training and Validation

We used both the traditional ML and modern DL approaches to build classification models and compared their performances for the problem of classifying a fish sample as safe or unsafe based on its histamine content, hopefully reflected by its extracted NIR spectral features. For the traditional ML approach, four algorithms were evaluated, including decision tree (DT) [28], k-nearest neighbors (KNN) [29], support vector machine (SVM) [30], and extreme gradient boosting (XGB) [31]. For the DL approach, we employed a convolutional neural network (CNN) [32] and proposed suitable architectures depending on the experiments.

As model hyperparameters can dramatically influence the performance of the conventional ML and CNN algorithms, hyperparameter tuning procedures on the common validation set were carried out to produce optimal models. Table 2 lists the set of hyperparameters used in the grid searching for the optimal ML models. Meanwhile, the process of optimization for CNN models involves adjusting two key components: the hyperparameters and the layers. While tuning the latter proves to be more challenging compared to the former, the former shares similarities with conventional ML algorithms. In the context of CNN models, the hyperparameters subject to tuning encompass the number of neurons, activation function, optimizer, learning rate, batch size, and epochs. The subsequent step involves fine-tuning the number of layers, a characteristic absent in other conventional ML algorithms. The number of layers employed in a CNN can significantly impact its accuracy. Insufficient layering may yield an underfitting outcome, whereas an excessive number of layers can lead to overfitting. The model training and hyperparameter tuning processes were conducted by using the scikit-learn toolkit for the conventional ML algorithms and the Keras framework for the CNN models. After the optimal models were determined, their performances were evaluated on the common test set, which will be reported in the next section.

**Table 2.** Set of hyperparameters used in the grid searching for the optimal traditional ML models.

| Model | Set of Hyperparameters  |
|-------|---|
| DT    | maximum depth of the tree, minimum number of samples at a leaf node |
| KNN   | number of neighbors   |
| SVM   | regularization parameter, kernel type                               |
| XGB   | number of decision trees, maximum depth of a tree, learning rate    |

### 3. Results

The initial analysis conducted, which is described in the preceding section, exhibited discernible variations in the absorbance values observed across distinct anatomical regions of the fish. Consequently, the present inquiry centers on the accurate identification of the safety classification pertaining to a given fish sample, specifically concerning the optimal approach to measuring the NIR spectrum. Is it more suitable to measure the spectrum at a specific predetermined location or at any of the eight designated positions of the fish? In order to address this query, a pair of experimental investigations were undertaken. In the first experiment, the analysis focused solely on NIR data obtained from a predetermined location on the fish's body (hereafter referred to as "position-dependent"). Consequently, sub-datasets were utilized, with each sub-dataset comprising only one-eighth of the complete dataset in terms of sample size. In contrast, the second experiment employed the whole NIR dataset obtained from all of the eight measurement positions (called "position-independent"). The primary objective of these experiments was to establish a highly effective model for the purpose of detecting the safety label associated with the fish sample.

#### 3.1. Traditional Machine Learning Models for Histamine Content Classification

Table 3 presents the accuracy of the optimally tuned DT models when being evaluated on the test sets according to measurement positions and feature types. It can be seen that the DT classifier achieved the highest accuracy of 87.2% when "Internal, tail" was used as the measurement position and "der1" was chosen as the feature vector. For the position-independent experiment, it obtained the highest accuracy of 77.2% when the feature type "prep + der1" was selected.

**Table 3.** Accuracy (%) on test sets of optimally tuned DT models.

| Position              | Feature Type |      |             |      |             |             |
|-----------------------|--------------|------|-------------|------|-------------|-------------|
|                       | orig         | prep | der1        | der2 | prep + der1 | prep + der2 |
| Skin, nape            | 71.7         | 79.7 | 81.7        | 81.4 | 79.3        | 79.3        |
| Skin, back            | 74.4         | 74.7 | 77.2        | 76.5 | 80.3        | 79.6        |
| Skin, tail            | 71.4         | 81.4 | 77.6        | 76.6 | 72.8        | 79.3        |
| Skin, stomach         | 68.6         | 72.4 | 79.0        | 77.2 | 81.4        | 77.9        |
| Internal, nape        | 73.1         | 79.3 | 77.9        | 72.1 | 77.2        | 76.2        |
| Internal, back        | 72.1         | 74.1 | 74.1        | 73.8 | 68.6        | 72.4        |
| <b>Internal, tail</b> | 72.4         | 76.2 | <b>87.2</b> | 83.4 | 86.6        | 82.1        |
| Internal, stomach     | 67.2         | 77.2 | 76.2        | 74.5 | 77.6        | 78.3        |
| <b>All positions</b>  | 73.5         | 77.1 | 73.7        | 73.7 | <b>77.2</b> | 74.4        |

Best accuracy scores together with the corresponding positions are in bold.

Table 4 presents the accuracy of the optimally tuned KNN models when being evaluated on the test sets according to measurement positions and feature types. It can be seen that the KNN classifier achieved the highest accuracy of 83.2% when "Internal, tail" was used as the measurement position and "prep + der1" was chosen as the feature vector. For the position-independent experiment, it obtained the highest accuracy of 78.8% when the feature type "prep + der1" was also selected.

**Table 4.** Accuracy (%) on test sets of optimally tuned KNN models.

| Position              | Feature Type |      |      |      |             |             |
|-----------------------|--------------|------|------|------|-------------|-------------|
|                       | orig         | prep | der1 | der2 | prep + der1 | prep + der2 |
| Skin, nape            | 82.3         | 74.1 | 76.2 | 72.8 | 75.2        | 75.2        |
| Skin, back            | 64.7         | 81.0 | 82.4 | 73.0 | 82.7        | 81.3        |
| Skin, tail            | 78.3         | 76.6 | 80.0 | 73.1 | 77.9        | 77.6        |
| Skin, stomach         | 79.7         | 76.2 | 84.8 | 80.7 | 75.2        | 75.5        |
| Internal, nape        | 72.1         | 75.2 | 70.3 | 69.7 | 75.5        | 72.8        |
| Internal, back        | 72.1         | 75.2 | 72.8 | 68.6 | 71.4        | 72.4        |
| <b>Internal, tail</b> | 83.1         | 81.4 | 81.4 | 77.6 | <b>83.2</b> | 81.7        |
| Internal, stomach     | 69.7         | 80.3 | 76.9 | 72.1 | 79.3        | 80.0        |
| <b>All positions</b>  | 76.0         | 79.1 | 75.1 | 71.8 | <b>78.8</b> | 78.2        |

Best accuracy scores together with the corresponding positions are in bold.

Table 5 presents the accuracy of the optimally tuned SVM models when being evaluated on the test sets according to measurement positions and feature types. It can be seen that the SVM classifier achieved the highest accuracy of 86.3% when “Internal, tail” was used as the measurement position and “prep” was chosen as the feature vector. For the position-independent experiment, it obtained the highest accuracy of 77.2% when the feature type “orig” was selected.

**Table 5.** Accuracy (%) on test sets of optimally tuned SVM models.

| Position              | Feature Type |             |      |      |             |             |
|-----------------------|--------------|-------------|------|------|-------------|-------------|
|                       | orig         | prep        | der1 | der2 | prep + der1 | prep + der2 |
| Skin, nape            | 78.3         | 85.9        | 73.8 | 74.1 | 82.9        | 84.6        |
| Skin, back            | 84.4         | 85.5        | 81.0 | 74.0 | 85.1        | 85.5        |
| Skin, tail            | 81.7         | 80.7        | 84.1 | 74.1 | 80.3        | 80.0        |
| Skin, stomach         | 77.2         | 82.4        | 74.1 | 74.1 | 82.4        | 82.1        |
| Internal, nape        | 83.1         | 80.7        | 74.1 | 74.1 | 79.7        | 80.3        |
| Internal, back        | 77.9         | 75.2        | 75.5 | 74.1 | 75.5        | 75.2        |
| <b>Internal, tail</b> | 85.2         | <b>86.3</b> | 84.1 | 74.1 | 85.3        | 85.7        |
| Internal, stomach     | 82.8         | 80.3        | 71.7 | 74.1 | 79.7        | 80.0        |
| <b>All positions</b>  | <b>77.2</b>  | 77.0        | 73.9 | 70.6 | 77.0        | 77.0        |

Best accuracy scores together with the corresponding positions are in bold.

Table 6 presents the accuracy of the optimally tuned XGB models when being evaluated on the test sets according to measurement positions and feature types. It can be seen that the XGB classifier achieved the highest accuracy of 90.3% when “Internal, tail” was used as the measurement position and “der2” was chosen as the feature vector. For the position-independent experiment, it obtained the highest accuracy of 83.6% when the feature type “prep + der1” was selected.

**Table 6.** Accuracy (%) on test sets of optimally tuned XGB models.

| Position              | Feature Type |      |      |             |             |             |
|-----------------------|--------------|------|------|-------------|-------------|-------------|
|                       | orig         | prep | der1 | der2        | prep + der1 | prep + der2 |
| Skin, nape            | 83.1         | 79.3 | 85.9 | 83.1        | 80.3        | 82.8        |
| Skin, back            | 82.7         | 79.6 | 83.0 | 81.3        | 82.0        | 81.3        |
| Skin, tail            | 73.1         | 79.7 | 79.3 | 86.6        | 79.0        | 82.1        |
| Skin, stomach         | 80.3         | 80.0 | 84.1 | 84.1        | 82.4        | 83.8        |
| Internal, nape        | 77.9         | 80.7 | 82.1 | 80.0        | 84.1        | 81.7        |
| Internal, back        | 72.1         | 75.2 | 76.2 | 78.3        | 73.8        | 75.5        |
| <b>Internal, tail</b> | 74.5         | 84.8 | 89.7 | <b>90.3</b> | 86.9        | 86.9        |
| Internal, stomach     | 72.4         | 80.7 | 82.8 | 79.3        | 84.8        | 83.4        |
| <b>All positions</b>  | 79.1         | 80.5 | 82.5 | 80.4        | <b>83.6</b> | 81.8        |

Best accuracy scores together with the corresponding positions are in bold.



### 3.2. Convolutional Neural Network Model for Histamine Content Classification

Similar to traditional machine learning models, the classification performance of CNN models is contingent upon both the type of input feature vectors and the measurement position. Table 7 presents the accuracy of the optimally tuned CNN models when being evaluated on the test sets according to measurement positions and feature types. It can be seen that the CNN classifier gained the highest accuracy of 93.1% when “Internal, tail” was used as the measurement position and “prep + der2” was chosen as the feature vector. For the position-independent experiment, it attained the highest accuracy of 81.0% when the feature type “der1” was selected.

**Table 7.** Accuracy (%) on test sets of optimally tuned CNN models.

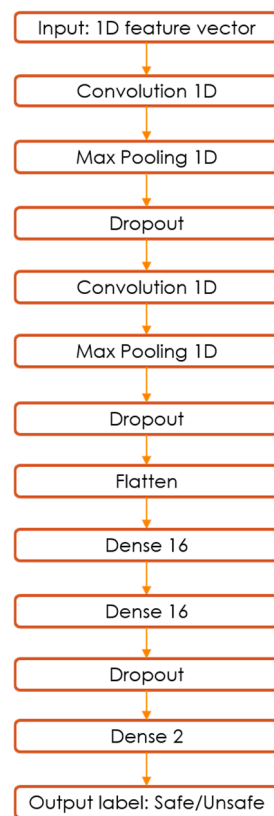
| Position              | Feature Type |      |             |      |             |               |
|-----------------------|--------------|------|-------------|------|-------------|---------------|
|                       | orig         | prep | der1        | der2 | prep + der1 | prep + der2   |
| Skin, nape            | 83.7         | 84.5 | 87.7        | 83.5 | 86.8        | 86.8          |
| Skin, back            | 81.0         | 81.2 | 81.2        | 74.4 | 80.5        | 82.5          |
| Skin, tail            | 82.6         | 89.1 | 88.4        | 85.1 | 87.5        | 86.2          |
| Skin, stomach         | 87.5         | 85.3 | 86.0        | 86.2 | 87.5        | 86.0          |
| Internal, nape        | 80.0         | 86.4 | 86.0        | 80.0 | 83.5        | 82.6          |
| Internal, back        | 84.8         | 81.7 | 81.3        | 76.6 | 81.5        | 83.3          |
| <b>Internal, tail</b> | 90.4         | 89.7 | 91.0        | 86.8 | 90.6        | <b>93.1 *</b> |
| Internal, stomach     | 87.1         | 87.5 | 89.7        | 84.0 | 87.7        | 90.0          |
| <b>All positions</b>  | 78.3         | 78.9 | <b>81.0</b> | 79.7 | 79.6        | 79.7          |

Best accuracy scores together with the corresponding positions are in bold. The highest classification accuracy among CNN models is marked with an asterisk.

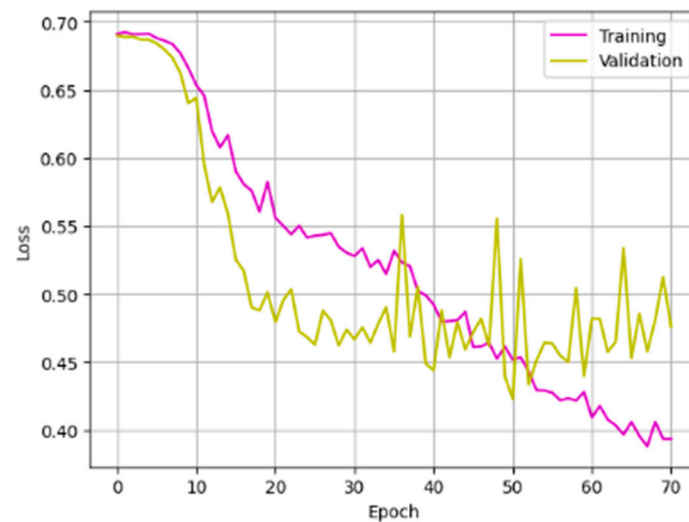
As each combination of the input feature type and the measurement position (and thus the corresponding NIR sub-dataset) leads to a different configuration of CNN model, we only present the process of constructing and evaluating the CNN model which achieved the highest classification accuracy (marked with an asterisk in Table 7) to make this article concise. Figure 6 describes the proposed CNN architecture in this case. The model includes one input layer which contains 456 neurons as input data, representing the feature vector of size  $456 \times 1$ , which is of the type “prep + der2” (i.e., pre-processed spectrum concatenated with its second derivative). It consists of two convolutional layers, each of them followed by a pooling layer and a dropout layer. The convolutional layers have kernels of size  $16 \times 1$  and Rectified Linear Units (ReLUs) as the activation functions. They are alternated with two max pooling layers with the pool size  $2 \times 1$  and two dropout layers with a rate of 0.01. The output of the final max pooling layer is streamed through a flatten layer in order to convert multi-dimensional data into one-dimensional data, which are then entered into the three fully connected (i.e., dense) layers. Both of the first two dense layers consist of 16 neurons and a ReLU activation function. A dropout layer is placed before the last dense layer. Finally, the last dense layer contains two neurons where softmax classifier activation is used to predict the output (i.e., the safety label) of the model. The proposed CNN model consists of 30,098 parameters.

The training process of this model was implemented using the Keras framework with the Adam optimizer and the initial learning rate at 0.0001. The learning rate was set to be reduced by a factor of 0.8 when the training result was not progressing. The validation set was used for stopping the training process. Given the substantial parameter count associated with the initial CNN architecture and the limited availability of training samples, the issue of overfitting emerged as a significant concern. Consequently, the incorporation of three dropout layers was deemed necessary in order to mitigate this challenge effectively.

Figure 7 shows how the cross-entropy-based loss function of the CNN model varied on the training and validation sets over training epochs. We stopped the training process after 70 epochs to prevent overfitting since the model had their losses converged on the validation set at this point.



**Figure 6.** The proposed CNN architecture which attains the best classification rate.



**Figure 7.** Loss variations on training and validation sets of the proposed CNN over training epochs.

In order to comprehensively assess the efficacy of the proposed CNN model in addressing the binary classification problem with imbalanced data, the following metrics were employed as evaluation measures on the test set:

- Accuracy =  $\frac{TP+TN}{TP+FP+TN+FN} = 93.1\%$ ;
- Sensitivity (or recall) =  $\frac{TP}{TP+FN} = 93.1\%$ ;
- Specificity =  $\frac{TN}{TN+FP} = 93.2\%$ ;

where

- TP (True Positive): The total number of samples where the model correctly predicts the positive class, i.e., when the actual class is Unsafe and the model also predicts it as Unsafe;
- TN (True Negative): The total number of samples where the model correctly predicts the negative class, i.e., when the actual class is Safe and the model also predicts it as Safe;
- FP (False Positive): The total number of samples where the model predicts the positive class incorrectly, i.e., the actual class is Safe but the model predicts it as Unsafe;
- FN (False Negative): The total number of samples where the model predicts the negative class incorrectly, i.e., the actual class is Unsafe but the model predicts it as Safe.

The evaluation of the proposed CNN model reveals not only a significantly high level of accuracy but also almost equally high specificity and sensitivity (or recall) values. This outcome establishes the model as highly effective and well suited for addressing the binary classification problem at hand.

Lastly, we evaluate the impact of the SMOTE technique on the performance of the proposed CNN model by trying not using SMOTE. Table 8 indicates sharp drops in accuracy and recall scores of the model when not using the SMOTE technique for balancing class distribution in the training set. This result confirms the essential use of SMOTE in dealing with a highly imbalanced dataset such as the one in our study.

**Table 8.** Performance of the proposed CNN model when using and not using SMOTE.

| Using SMOTE | Recall (%) | Accuracy (%) |
|-------------|------------|--------------|
| Yes         | 93.1       | 93.1         |
| No          | 74.4       | 80.3         |

### 3.3. Comparison among Different Classifiers

Table 9 summarizes the best performances of the investigated classifiers in the two experiments: position-independent and position-dependent. It can be observed that the position-dependent models attained a remarkably higher classification accuracy than the position-independent ones with the best position for NIR measurement being the “Internal, tail” part of the fish regardless of the classifiers. This suggests that we should collect NIR spectra inside the flesh at the tail for fish safety classification. Among the position-dependent models, the proposed CNN model described in Section 3.2 combined with the feature vector consisting of the pre-processed spectrum coupled with its second derivative was proved to be superior to the others. It achieved the highest classification accuracy of 93.1% on the hold-out test set and a similar level of specificity and recall scores.

**Table 9.** Summary of the best cases of different classifiers (accuracies for hold-out and cross-validation approaches are exhibited outside and inside parentheses, respectively).

| Classifier | Position-Independent |             | Position-Dependent |                |             |
|------------|----------------------|-------------|--------------------|----------------|-------------|
|            | Accuracy (%)         | Feature     | Accuracy (%)       | Position       | Feature     |
| DT         | 77.2 (78.6)          | prep + der1 | 87.2 (86.9)        | Internal, tail | der1        |
| KNN        | 78.8 (77.5)          | prep + der1 | 83.2 (84.4)        | Internal, tail | prep + der1 |
| SVM        | 77.2 (76.3)          | orig        | 86.3 (85.4)        | Internal, tail | prep        |
| XGB        | 83.6 (82.8)          | prep + der1 | 90.3 (89.8)        | Internal, tail | der2        |
| <b>CNN</b> | 81.0 (80.4)          | der1        | <b>93.1 (92.7)</b> | Internal, tail | prep + der2 |

Best accuracy scores together with the corresponding classifier are in bold.

To ensure the reliability of the reported experiment results, we further employed a stratified five-fold cross-validation technique on the NIR dataset for re-evaluating the optimal classifiers. We maintained the same optimal hyperparameters as found in the previous experiments and used the entire dataset for model training and testing. The

stratified five-fold cross-validation strategy included splitting the dataset into an 80% training set and a 20% test set for each fold so as to ensure the proportion of the Safe/Unsafe samples was the same across the training set and the test set, which gives a more accurate estimate of classification performance. In each fold, the training set was pre-processed as described in Section 2.3 before a model was fitted and then evaluated on the test set. Finally, the mean of the model's accuracy after running five folds was used to provide its cross-validation performance, which is shown by the numbers in parentheses in Table 9. It can be seen that there are insignificant differences between hold-out and cross-validation results. This could be accounted for by the careful training/validation/test data separation we carried out when using the hold-out approach. These results confirm the robustness of our findings in this section.

### 3.4. Discussion on the Optimal Measurement Point

Histamine is formed by the activity of microbial enzymes. We predict that microorganisms are more active on the surface than inside the fish, leading to a higher abundance of histamine. Additionally, the distribution of histidine, the substrate for histamine formation, is not uniform within the fish meat. Therefore, the histamine content at our measurement points may vary. According to Vietnamese standards and other guidelines, safety related to histamine content is determined based on the average histamine content of the entire fish. Thus, in our study, classification is based on the relationship between the NIR signal reflection on the surface of the measured sample and the average histamine content of the entire fish. The finding that the tail of the fish provides the best prediction accuracy emerged from data processing and modeling. We believe that the histamine concentration at this location can approximate the average histamine content of the entire fish when thoroughly mixed.

## 4. Conclusions

In this study, we employed NIR spectroscopy in conjunction with ML techniques, including DL using CNNs, to classify fish samples into two safety classes, Safe and Unsafe, based on their histamine content. This study developed an effective machine learning workflow that addresses class imbalance in the collected data by incorporating the SMOTE technique and investigating various combinations of feature extraction techniques and ML/DL algorithms. The main findings of this study can be summarized as follows. Firstly, utilizing the NIR dataset collected at a specific location within the fish, specifically inside the flesh at the tail, yielded superior models for fish safety classification. Secondly, a feature extraction technique based on the original NIR spectrum, coupled with a CNN architecture optimized for the task, outperformed conventional ML classifiers, achieving an accuracy of approximately 93%. These findings have significant potential in developing a fast and cost-effective method for detecting fish safety, providing valuable insights for food safety authorities for determining the need for more advanced and expensive experiments in assessing fish quality in laboratory settings.

It is important to note that our research was constrained by a small-sized NIR dataset obtained using a low-cost NIR scanner and a limited number of fish samples from three specific types: mackerel, tuna, and pompano. To enhance the study's scope and robustness, future research will involve expanding the dataset by employing higher quality NIR scanners to train more powerful classification models. Additionally, conducting similar studies on the detection of other potentially harmful agents in fish, such as urea and borax, would help validate the effectiveness of NIR spectroscopy and the suggested ML workflow. Lastly, considering the incorporation of chemometrics and multivariate data analysis methods such as Principal Component Analysis or Partial Least Squares Discriminant Analysis could be beneficial for extracting more relevant features from NIR spectra, enabling differentiation among multiple safety classes.

**Author Contributions:** Conceptualization, D.K.N.; methodology, D.K.N. and K.D.P.; software, K.D.P.; validation, D.K.N. and M.N.D.; investigation, D.K.N. and C.T.V.; data curation, D.K.N., C.T.V. and M.N.D.; writing—original draft preparation, D.K.N. and M.N.D.; writing—review and editing, D.K.N. and K.D.P.; supervision, N.L.T.; project administration, N.L.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministry of Science and Technology of Vietnam under project number ĐTĐL.CN-33/20.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** This work was supported by the Ministry of Science and Technology of Vietnam in the project “Application of quick analysis methods combining multi-dimensional data processing and machine learning in quality control of some types of seafood” (Project No.: ĐTĐL.CN-33/20).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Visciano, P.; Schirone, M.; Tofalo, R.; Suzzi, G. Histamine poisoning and control measures in fish and fishery products. *Front. Microbiol.* **2014**, *5*, 500. [CrossRef] [PubMed]
2. U.S. Department of Health and Human Services; Food and Drug Administration; Center for Food Safety and Applied Nutrition; Office of Food Safety. *Fish and Fishery Products Hazards and Controls Guidance*, 4th ed.; U.S. Department of Health and Human Services: Washington, DC, USA, 2011; pp. 113–152.
3. Vietnamese Standard on Tuna’s Raw Material (In Vietnamese). Available online: <https://tieuchuan.vsqi.gov.vn/tieuchuan/view?sohieu=TCVN+12153:2018> (accessed on 10 August 2024).
4. Surya, T.; Sivaraman, B.; Alamelu, V.; Priyatharshini, A.; Prabu, E.; Sundhar, S. Rapid methods for histamine detection in fishery products. *Int. J. Curr. Microbiol. Appl. Sci.* **2019**, *8*, 2035–2046. [CrossRef]
5. Wenqian, Y. Applications of near infrared spectroscopy for fish and fish products quality: A review. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *657*, 012115.
6. Ding, R.; Huang, X.; Han, F.; Dai, H.; Teye, E.; Xu, F. Rapid and nondestructive evaluation of fish freshness by near infrared reflectance spectroscopy combined with chemometrics analysis. *Anal. Methods* **2014**, *6*, 9675–9683. [CrossRef]
7. Kimiya, T.; Sivertsen, A.H.; Heia, K. VIS/NIR spectroscopy for non-destructive freshness assessment of Atlantic salmon (*Salmo salar* L.) fillets. *J. Food Eng.* **2013**, *116*, 758–764. [CrossRef]
8. Shim, K.; Jeong, Y. Freshness evaluation in chub mackerel (*Scomber japonicus*) using near-infrared spectroscopy determination of the cadaverine content. *J. Food Prot.* **2019**, *82*, 768–774. [CrossRef]
9. Sivertsen, A.H.; Kimiya, T.; Heia, K. Automatic freshness assessment of cod (*Gadus morhua*) fillets by Vis/Nir spectroscopy. *J. Food Eng.* **2011**, *103*, 317–323. [CrossRef]
10. Zhou, J.J.; Wu, X.Y.; Chen, Z.; You, J.; Xiong, S.B. Evaluation of freshness in freshwater fish based on near infrared reflectance spectroscopy and chemometrics. *LWT-Food Sci. Technol.* **2019**, *106*, 145–150. [CrossRef]
11. Isaksson, T.; Tøgersen, G.; Iversen, A.; Hildrum, K.I. Non-destructive determination of fat, moisture and protein in salmon fillets by use of near-infrared diffuse spectroscopy. *J. Sci. Food Agric.* **1995**, *69*, 95–100. [CrossRef]
12. Khodabux, K.; L’Omelette, M.S.S.; Jhaumeer-Laulloo, S.; Ramasami, P.; Rondeau, P. Chemical and near-infrared determination of moisture, fat and protein in tuna fishes. *Food Chem.* **2007**, *102*, 669–675. [CrossRef]
13. Wold, J.P.; Isaksson, T. Non-destructive determination of fat and moisture in whole atlantic salmon by near-infrared diffuse spectroscopy. *J. Food Sci.* **1997**, *62*, 734–736. [CrossRef]
14. Lv, H.; Xu, W.; You, J.; Xiong, S. Classification of freshwater fish species by linear discriminant analysis based on near infrared reflectance spectroscopy. *J. Near Infrared Spectrosc.* **2017**, *25*, 54–62. [CrossRef]
15. Cozzolino, D.; Chree, A.; Scaife, J.R.; Murray, I. Usefulness of near-infrared reflectance (NIR) spectroscopy and chemometrics to discriminate fishmeal batches made with different fish species. *J. Agric. Food Chem.* **2005**, *53*, 4459–4463. [CrossRef]
16. López-Maestresalas, A.; Insausti, K.; Jarén, C.; Pérez-Roncal, C.; Urrutia, O.; Beriain, M.J.; Arazuri, S. Detection of minced lamb and beef fraud using NIR spectroscopy. *Food Control* **2019**, *98*, 465–473. [CrossRef]
17. Pereira, E.V.D.S.; Fernandes, D.D.D.S.; de Araújo, M.C.U.; Diniz, P.H.G.D.; Maciel, M.I.S. Simultaneous determination of goat milk adulteration with cow milk and their fat and protein contents using NIR spectroscopy and PLS algorithms. *LWT* **2020**, *127*, 109427. [CrossRef]
18. Varrà, M.O.; Fasolato, L.; Serva, L.; Ghidini, S.; Novelli, E.; Zanardi, E. Use of near infrared spectroscopy coupled with chemometrics for fast detection of irradiated dry fermented sausages. *Food Control* **2020**, *110*, 107009. [CrossRef]
19. Barbin, D.F.; ElMasry, G.; Sun, D.-W.; Allen, P.; Morsy, N. Non-destructive assessment of microbial contamination in porcine meat using NIR hyperspectral imaging. *Innov. Food Sci. Emerg. Technol.* **2013**, *17*, 180–191. [CrossRef]
20. Puertas, G.; Vázquez, M. Fraud detection in hen housing system declared on the eggs’ label: An accuracy method based on UV-VIS-NIR spectroscopy and chemometrics. *Food Chem.* **2019**, *288*, 8–14. [CrossRef] [PubMed]

21. Maione, C.; Barbosa, F.; Barbosa, R.M. Predicting the botanical and geographical origin of honey with multivariate data analysis and machine learning techniques: A review. *Comput. Electron. Agric.* **2019**, *157*, 436–446. [[CrossRef](#)]
22. Gupta, O.; Das, A.J.; Hellerstein, J.; Raskar, R. Machine learning approaches for large scale classification of produce. *Sci. Rep.* **2018**, *8*, 5226. [[CrossRef](#)]
23. Guo, J.; Chen, C.; Zuo, E.; Dong, B.; Lv, X.; Yang, W. Near-infrared spectroscopy combined with pattern recognition algorithms to quickly classify raisins. *Sci. Rep.* **2022**, *12*, 7928. [[CrossRef](#)] [[PubMed](#)]
24. Benmouna, B.; García-Mateos, G.; Sabzi, S.; Fernandez-Beltran, R.; Parras-Burgos, D.; Molina-Martínez, J.M. Convolutional neural networks for estimating the ripening state of Fuji apples using visible and near-infrared spectroscopy. *Food Bioprocess. Technol.* **2022**, *15*, 2226–2236. [[CrossRef](#)]
25. Stroka, J.; Bouten, K.; Mischke, C.; Breidbach, A.; Ulberth, F. *Equivalence Testing of Histamine Methods—Final Report*; Publications Office of the European Union: Luxembourg, 2014; ISBN 9789279378003.
26. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
27. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [[CrossRef](#)]
28. Quinlan, J.R. Simplifying decision trees. *Int. J. Man-Mach. Stud.* **1987**, *27*, 221–234. [[CrossRef](#)]
29. Cover, T.M.; Hart, P.E. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [[CrossRef](#)]
30. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
31. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794.
32. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems—Volume 1 (NIPS'12)*, Lake Tahoe, NV, USA, 3–6 December 2012; Curran Associates Inc.: Red Hook, NY, USA, 2012; pp. 1097–1105.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.