big data and cognitive computing

Article

Classification of Microbiome Data from Type 2 Diabetes Mellitus Individuals with Deep Learning Image Recognition

Juliane Pfeil [1], Julienne Siptroth [2], Heike Pospisil [2], Marcus Frohme [1,*], Frank T. Hufert [3], Olga Moskalenko [4], Murad Yateem [4] and Alina Nechyporenko [1,5]

[1] Division Molecular Biotechnology and Functional Genomics, Technical University of Applied Sciences, 15745 Wildau, Germany
[2] Division High Performance Computing in Life Sciences, Technical University of Applied Sciences, 15745 Wildau, Germany
[3] Institute for Microbiology and Virology, Brandenburg Medical School Theodor Fontane, 16816 Neuruppin, Germany
[4] BIOMES NGS GmbH, 15745 Wildau, Germany
[5] Department of Systems Engineering, Kharkiv National University of Radio Electronics, 61166 Kharkiv, Ukraine
* Correspondence: mfrohme@th-wildau.de; Tel.: +49-(0)-3375-508-249

Abstract: Microbiomic analysis of human gut samples is a beneficial tool to examine the general well-being and various health conditions. The balance of the intestinal flora is important to prevent chronic gut infections and adiposity, as well as pathological alterations connected to various diseases. The evaluation of microbiome data based on next-generation sequencing (NGS) is complex and their interpretation is often challenging and can be ambiguous. Therefore, we developed an innovative approach for the examination and classification of microbiomic data into healthy and diseased by visualizing the data as a radial heatmap in order to apply deep learning (DL) image classification. The differentiation between 674 healthy and 272 type 2 diabetes mellitus (T2D) samples was chosen as a proof of concept. The residual network with 50 layers (ResNet-50) image classification model was trained and optimized, providing discrimination with 96% accuracy. Samples from healthy persons were detected with a specificity of 97% and those from T2D individuals with a sensitivity of 92%. Image classification using DL of NGS microbiome data enables precise discrimination between healthy and diabetic individuals. In the future, this tool could enable classification of different diseases and imbalances of the gut microbiome and their causative genera.

Keywords: human intestinal microbiome; next-generation sequencing; type 2 diabetes; deep learning; image classification

check for updates

Citation: Pfeil, J.; Siptroth, J.; Pospisil, H.; Frohme, M.; Hufert, F.T.; Moskalenko, O.; Yateem, M.; Nechyporenko, A. Classification of Microbiome Data from Type 2 Diabetes Mellitus Individuals with Deep Learning Image Recognition. *Big Data Cogn. Comput.* **2023**, *7*, 51. https://doi.org/10.3390/bdcc7010051

Academic Editors: Robail Yasrab and Md Mostafa Kamal Sarker

Received: 15 February 2023
Revised: 2 March 2023
Accepted: 15 March 2023
Published: 17 March 2023

1. Introduction

The analysis of the human microbiome is an innovative field of research, which in particular investigates the interaction of the intestinal flora and aims to draw conclusions about the general state of health and causes of diseases. The microbiome is composed of all microorganisms that colonize multicellular organisms. These are also termed microbiota and consist of a variety of interacting bacteria, viruses, and fungi. In this context, the term dysbiosis is associated with an imbalance in the intestinal flora, which can have various reasons [1].

Research has shown that alterations of the gut microbiome or the presence of certain pathologically relevant species can be associated with vitamin deficiency [2], obesity [3], inflammatory bowel diseases [4] and colon cancer [5], and autoimmune [6] and neurodegenerative disorders [7]. The microbiome is an intensively researched field, but the relevance of many factors still remains to be fully explained and it is uncertain to what extent the intestinal flora is influenced by genetic factors or environmental conditions (diet, sport, etc.) [8]. To highlight the importance of this area of research, the European cooperation in science and technology (COST) action ML4microbiome (statistical and machine learning techniques in human microbiome studies) was launched, bringing together microbiome researchers and experts in the field of machine learning (ML). The objective is to optimize and standardize analytical methods and to provide publicly available benchmark datasets [9].

In general, a fast and accurate classification of a disease is essential for its treatment, and early diagnoses improve the chances of recovery and can help to minimize consequential damage. Furthermore, this knowledge promotes a better understanding of a disease and can help to assess its evolution. This can form the basis for the development of effective drugs or treatment methods and a possible "early warning system" for diseases.

In the case of type 2 diabetes mellitus (T2D), special attention is paid to pre-diabetes—a precursor of diabetes mellitus disease in which glycemic parameters are already elevated, but the threshold for T2D has not yet been reached [10]. Early diagnosis and therapy can delay or prevent the development of possible secondary diseases such as diabetic foot syndrome, diabetic retinopathy, or diabetic neuropathy; all are severe and potentially lead to amputation, blindness, or even death [11].

Next-generation sequencing (NGS) technology can identify and quantify the majority of all bacterial and fungal species from a stool sample [12]. Herein the analysis of the obtained sequence reads is performed in several steps until biologically normalized counts per taxonomic level are received. For this purpose, the calculation of the abundances can be performed using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2) [13]. Multi-omics data are known to be complex and heterogeneous, requiring the application of advanced dimensionality reduction techniques. Regarding the microbiome, it is a dynamic ecosystem with active host regulation [14], which significantly increases the complexity in the analysis and interpretation of data. For example, missing values and large numbers of sparse values have to be handled [15]. One way to address these challenges is to transform the data into a different format of representation. The creation of a suitable depiction reduces the complexity and simplifies the analysis situation by making it more compact. This opens up the possibility of implementing other approaches and methods. For example, visualized data representations can be processed and analyzed using methods from the field of computer vision. Limitations of tabular data can thus be overcome [15].

For the analysis of complex and large datasets, ML/deep learning (DL) has proven to be particularly advantageous. Different ML techniques, such as clustering and dimensional-reduction-based approaches, random forest (RF), regression models, and support vector machines (SVMs) have been applied to analyze microbiome sequence data. These methods have been used to investigate the microbial community and their influence on different phenotypes. Recent research studies are mainly concerned with the identification of disease-related profiles and risk-prediction biomarkers [16]. For the detection of T2D, regression models, Bayes classifiers, RF, and SVM were used [17,18]. Based on operational taxonomic units (OTUs), these conventional classification methods achieve a maximum area under the curve (AUC) of 0.74 with RF [17] and a sensitivity of 75% and specificity of 69% (AUC = 0.76) with a regression model [18]. DL techniques as recurrent neural networks were used to investigate temporal dependencies from long-term datasets for the prediction of food allergies, nutrition, and diseases. Autoencoders were applied for dimensionality reduction to create latent representations that improve prediction accuracies [19]. Researchers classified colorectal cancer (CRC) using the OTU table and deep neural networks with a sensitivity of 88% and a specificity of 98% [20]. DL image recognition can also contribute to the diagnosis, and the application of a residual network (ResNet) [21] to microscopic images showed an accuracy of 80%, a sensitivity of 87%, and a specificity of 83% [22]. In the field of image recognition, DL convolutional neural networks (CNNs) are characterized by accuracy, robust deployment, and generalization potential [23,24]. Furthermore, several research efforts are attempting to apply the potential and advantages of these methods to non-image data. Gene expression and microbiome sequence data have been visualized using different methods, and CNNs have been used to classify phenotypes and diseases. Reiman et al. generated phylogenetic and taxonomic trees of microbiome data and transformed them into matrices [25,26]. These could be classified as images to predict phenotypes of origin such as skin, mouth, and gut. To classify T2D, liver cirrhosis (LC), obesity, and CRC, researchers have used taxonomic representations [27], OTU clustering [17], colormaps [28], and phylogenetic trees [29]. The CNNs outperform conventional ML techniques such as RF or SVM, but still achieve moderate values for T2D prediction, such as maximum AUC values of 0.81 [27], 0.67 [29], and 0.75 [17], and an accuracy of 0.68 [28]. Other researchers transformed gene expression data into feature matrices or heatmaps to predict cancer, lymphoma, and Parkinson's disease [30,31].

State-of-the-art DL models comprise the family of ResNets [21] that have been used for many problems and datasets (e.g., ImageNet [32], Modified National Institute of Standards and Technology (MNIST) database [33], and Canadian Institute For Advanced Research (CIFAR) dataset [34]). ResNet models solved the degradation problem originating from vanishing gradients by the introduction of residual blocks that allow skipping of certain layers and contain non-linearities (ReLU—rectifier linear units). Thereby, ResNet offers good performance with a large number of layers and reasonable training efforts. Furthermore, in comparison to other network architectures (e.g., GoogLeNet [35] or visual geometry group (VGG) [36]), ResNet generates a higher classification accuracy [24] and can also serve as a backbone for advanced image recognition tasks (e.g., object detection [37,38] or image segmentation [39,40]). Research effort has been made to use the ResNet model for taxonomic representations of sequenced microbiome datasets to classify T2D, CRC, and LC. The drawbacks of this approach are overfitting and the detrimental influence of ImageNet pretrained weights [27]. Michel-Mata et al. tried to use ResNet to predict microbiome compositions from different communities, but their concept suffered from restrictions of the dataset [41]. Consequently, neither the visualization or DL CNN approaches are capable of predicting T2D from microbiome data with satisfying accuracy, sensitivity, and specificity; nor has ResNet been successfully used in this context.

An important consideration is that due to the lack of standards, the characteristics and quality of microbiome data in publicly available databases vary. This is caused not only by different experimental conditions and sample preparation factors, but also by different methods of data preprocessing, such as sequence filtering, clustering, and taxonomic assignment, as well as other methods and tools used in bioinformatic pipelines. Moreover, the integration of relevant metadata such as gender, age, nutrition, lifestyle, and other factors that are critical for obtaining meaningful information from microbiome studies is difficult due to the lack of detailed and structured metadata in public data resources [16].

The aim of this research work was to develop an appropriate visualization technique for microbiome sequence data that enables the use of DL image recognition to analyze their characteristics. For this purpose, the DL model ResNet was trained and optimized with the visualized data to enable accurate classification between healthy and sick individuals. In this context, the disease T2D was selected as a proof-of-concept. The results were evaluated by determining accuracy, specificity, and sensitivity. Different visualizations of the microbiome data at phylum, class, and genus levels were explored to assess robustness and generalization potential.

This work is structured as follows: Section 2 describes the NGS methods used to generate the microbiome dataset and its characteristics, the applied visualization techniques, and the parameters of the DL image classification model ResNet. Section 3 presents and evaluates the classification results for all visualizations. Section 4 discusses the results in the context of previous research and suggests approaches for further investigation.

2. Materials and Methods

2.1. Sample Preparation and NGS Data Processing

A detailed outline of the sample preparation, NGS, and processing of the sequence reads follows Siptroth et al. [42]. Stool samples from customers of BIOMES NGS GmbH served as a source for microbiome data and were collected from self-tests for the analysis of the intestinal flora. These tests are not a diagnostic product under medical surveillance. Only data for which the study participants have declared their agreement to scientific research were used. Samples were stored frozen until lysis and DNA extraction, and library preparation for sequencing on Illumina MiSeq was performed. Processing of the bacterial 16S ribosomal DNA sequence reads started by filtering of the determined paired-end reads followed by the clustering of the reference sequences according to their similarity using Cluster Database at High Identity with Tolerance (CD-HIT) [43,44]. The calculation of the biologically normalized counts was conducted using the PICRUSt2 pipeline.

2.2. Dataset and Study Group

The dataset contains more than 29,000 samples including microbiome profiles and associated individual lifestyle data that include information on age, body mass index (BMI), diet, and other characteristics (details in Siptroth et al. [42]). These lifestyle data are self-reported by the study participants in a questionnaire, the results of which are provided by BIOMES NGS GmbH. The microbiome profiles contain relative counts per taxonomic level (phylum to species). The inclusion criterion for all participants for this research was an age between 18 and 80 years. For classification into the selected groups 'healthy' and 'T2D' the main parameter was the self-report of a subject as a type 2 diabetes mellitus patient. In the healthy group, exclusion criteria were a BMI lower than 18.5 or higher than 27.5, or any known diseases, gastrointestinal complaints, gluten intolerances, medication, or probiotics intake within the last three months. Further exclusion criteria were daily alcohol consumption, a well-being score lower than 4 (out of 10), or a health score lower than 6 (out of 10). This narrowed down the number of eligible samples to 674 for the healthy group and 272 for the T2D group. Table 1 lists the age, sex, and BMI parameters for the two study groups.

Table 1. Distribution of the age, sex, and body mass index (BMI) parameters for the healthy and type 2 diabetes mellitus (T2D) groups.

Study Group	Age [Years]	Women/Men/Other	BMI [kg/m^2]
healthy	42.55 ± 12.12	340/318/16	23.13 ± 2.24
T2D	59.71 ± 12.27	143/127/2	31.05 ± 6.38

2.3. Visualization Methods

In order to perform image classification, the obtained relative counts of all microbial genera in the NGS data were transformed into another form of representation. Different visualization methods (t-distributed stochastic neighbor embedding (t-SNE), taxonomic trees and graphs, and histograms and stacked histograms) were evaluated. On the one hand the represented formats were basically not suitable for DL image classification because visualizations at the sample level were not useful (t-SNE), scaling was difficult due to very small and large values and non-uniform image sizes (histogram), or the resulting representations were just too wide (taxonomic trees and graphs). On the other hand, no classification success could be achieved (stacked histogram) mainly because of the lack of features. A heatmap visualization implemented by the Python library seaborn [45] was chosen to display the data, showing the relative counts at genus level. These were filtered by excluding all unspecific genera (no exact assignment could be made at this or a higher taxonomic level) and those present in less than 5% of all samples. Out of 2445 genera, 362 remained for visualization. These were sorted alphabetically by family level and gray-scales were used to represent the relative abundance of each genus. Logarithmic scaling of the color palette was performed to obtain a higher sensitivity towards small values. The gray-scale representation and the corresponding scaling were particularly advantageous in comparison to a color-based visualization. This approach showed the best contrast and revealed very small abundances that offered the most suitable representation for image classification with a large number of evaluable features. An image for each sample was generated as a 1D heatmap. An image of a sample of the T2D group and the respective legend for logarithmic scaling of abundances as gray-scales can be seen in Figure 1.

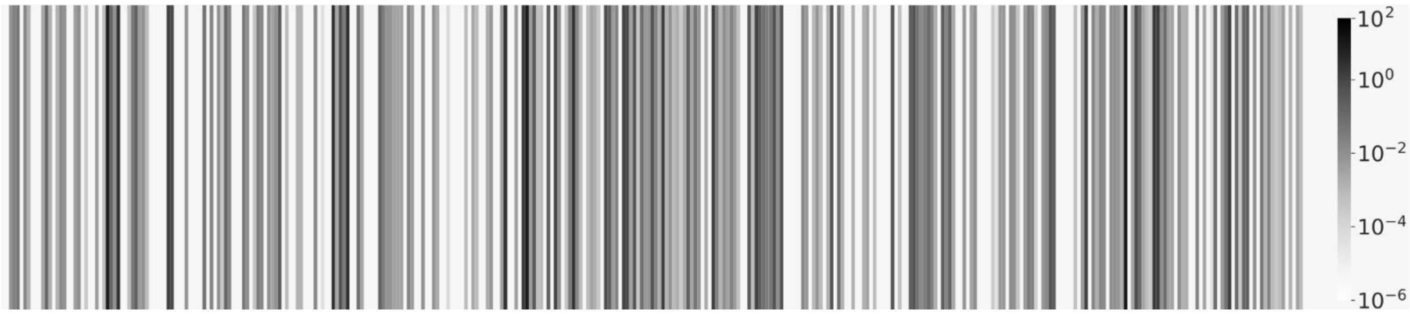

Figure 1. The relative counts of the genus levels of a T2D sample represented by a 1D heatmap. The legend shows the logarithmic scaling of abundances of the bacteria as gray-scales.

These 1D heatmaps were transformed further into a radial representation (using the polar projection of seaborn [45]) by extracting the intensities from left to right and plotting them starting at 12 o'clock, which represented 0°. All abundances were given the same circular area so that each of the respective 362 genera filled 0.99°. For the radial heatmap representation, a high resolution of 2000 × 2000 pixels (px) was chosen with the aim to compress the images and to convert them to a suitable format for DL image classification with easily distinguishable intensities (Figure 2). All other information (axis labels, legends, etc.) was removed.

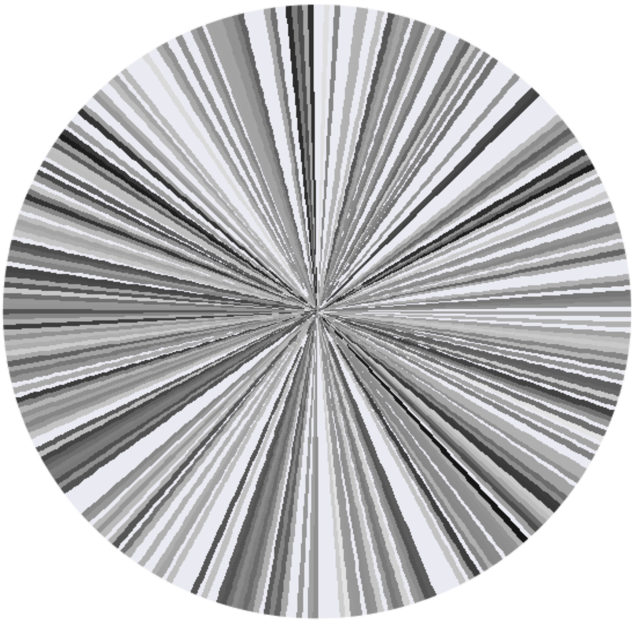

Figure 2. Radial representation showing the relative counts of the genus level of a T2D sample (2000 × 2000 px). The abundances of the bacteria are shown as logarithmically scaled gray levels.

The same approach was used to convert the relative counts of the phylum and class levels into radial heatmaps. For the phylum level, 62 phyla of 76, and for the class level 115 out of 228, remained after filtering.

To ensure the robustness of the visualization, the training was performed with different modifications of the original dataset at the genus level. The visualization was rotated, mirrored, and shuffled. Table 2 lists the modifications.

Table 2. Modifications of original dataset.

Dataset	Properties
original	genera sorted alphabetically by family level
90°	original dataset 90° rotated clockwise
180°	original dataset 180° rotated clockwise
270°	original dataset 270° rotated clockwise
vertical	original dataset vertical mirrored
horizontal	original dataset horizontal mirrored
shuffled_a	original dataset randomly shuffled
shuffled_b	original dataset randomly shuffled

A Monte Carlo cross validation (MCCV) with three repetitions [46] was performed with the ratio of 60:20:20 of these datasets resulting in three training, validation, and test sets. The training sets contained 566 images (404 healthy, 162 T2D), and the validation and test sets each contained 190 images (135 healthy, 55 T2D). Accordingly, for each modified dataset three individual models were trained. To evaluate the reliability of the trained classifiers, a dataset with mislabeled data was prepared: from the original dataset, 50% of T2D samples were labeled as ‘healthy’ and the corresponding amount of the healthy group was labeled as ‘T2D’.

2.4. ML/DL Algorithms and Training

For image classification, the well-known and powerful residual network with 50 layers (ResNet-50) model was selected [21]. Training, validation, and testing was conducted with the ResNet-50 keras implementation. Keras is a high-level application programming interface (API) and DL library that simplifies the usage of neural networks [47]. Training parameters for ResNet-50 have been customized for the existing graphics processing unit (GPU) infrastructure (GPU NVIDIA Tesla V100, DDR4-RAM 384 GB) and the task of image classification. In order to optimize the classification accuracy, the network was trained from scratch (no pretrained weights were used), all layers were set as trainable, and the settings shown in Table 3 were applied.

Table 3. Training parameters for ResNet-50.

Epoch Number	Loss	Batch Size	Image Size	Optimizer		
				Class	Learning Rate	Epsilon
100	categorical cross-entropy	4	512 × 512 px	Adam	0.001	10^{-8}

During the training process, resizing of the original images to 512 × 512 px proved to be particularly beneficial. A larger image size did not improve classification accuracy and increased training time significantly.

The results of the image classification models were compared with conventional ML techniques such as RF [48] and SVM [49]. For this purpose a 5-fold cross validation [46] was applied to 80% of the dataset and the remaining test data were used to evaluate the performance. A grid search approach determined the best hyperparameter set.

3. Results

The aim of this research was to provide a simple, fast, and accurate classification of visualized microbiome profiles of healthy persons and individuals with T2D using DL image recognition. Several models of the ResNet-50 neural network were trained with radial heatmap visualizations of phylum, class, and different arrangements of genera level including a MCCV with three repetitions. The results were evaluated in terms of accuracy, specificity, and sensitivity. By determining the true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), these parameters can be calculated using the following equations:

$$\text{accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{FN} + \text{TN}) = \text{all correct predicted samples}/\text{all samples} \quad (1)$$

$$\text{sensitivity} = \text{TP}/(\text{TP} + \text{FN}) = \text{all samples correctly predicted as diseased}/\text{all diseased samples} \quad (2)$$

$$\text{specificity} = \text{TN}/(\text{TN} + \text{FP}) = \text{all samples correctly predicted as healthy}/\text{all healthy samples} \quad (3)$$

During the training process, the models were optimized with the accuracy and loss metric. The accuracy expression basically describes which proportions of the two groups (healthy, T2D) were correctly classified on average. The loss value is calculated from the neural network objective function and represents the error to be minimized. Training with the phylum and class level visualizations was not successful. Accuracies of about 0.7 were achieved, resulting in only the healthy class being detected. Therefore, the data derived from the phylum and class level were not explored further. At genus level, the accuracy for the training set was always 1.0 after 100 training epochs; hence, the results for this set will not be investigated more closely. After approximately 50 epochs, the values for the validation sets were stable. However, to really ensure training success, 100 epochs were chosen for the entire training. No over-fitting, which would have been visible by a decrease in accuracies, was detected. In order to represent these investigations, the following figure shows the training and validation accuracy (Figure 3a) and loss (Figure 3b) that are represented for an example model of the original dataset at the genus level.

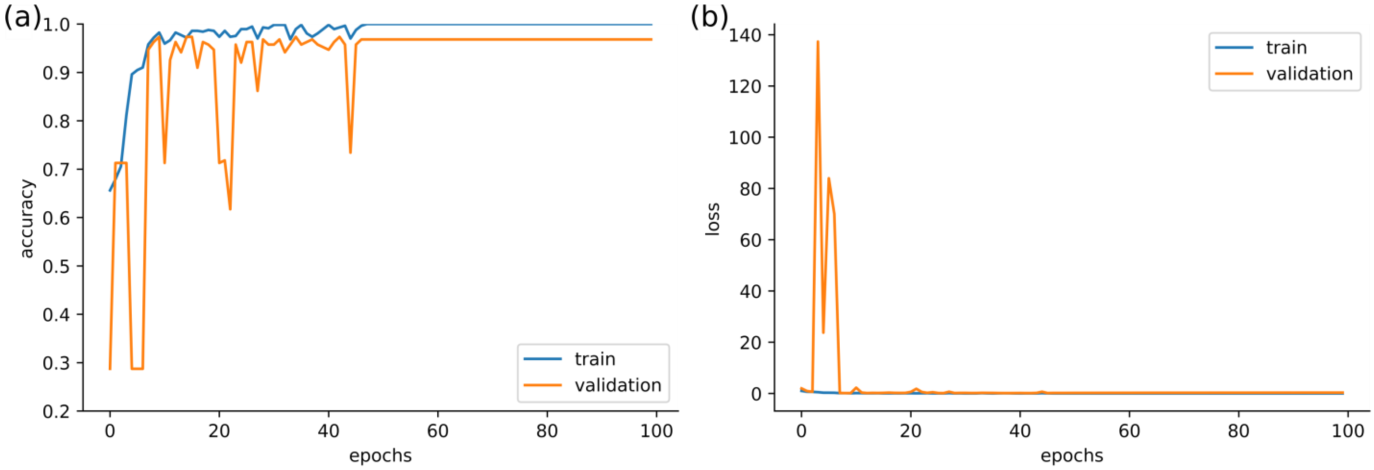

Figure 3. Accuracy and loss of the training and validation set from the original dataset. Results are presented for genus level data. Parameters for the training set are visualized as a blue line and for the validation set as an orange line. The values were tracked with keras [47]. (**a**) Accuracy of the training and validation set during the training of 100 epochs. (**b**) Training and validation set loss during the training of 100 epochs.

The display shows that for the training set, the accuracy immediately rises and reaches values around 1.0 after 10 epochs. The loss is described by small values in the beginning and does not fluctuate. For the validation set, accuracy and loss vary a lot during early epochs. After 10 epochs, the loss becomes very small, and the accuracy starts to increase and becomes stable around 50 epochs. For all other datasets, the accuracy and loss behave similarly, and it can be assumed that 50 epochs would be sufficient for training (a higher number of epochs has no disadvantages, because over-fitting could not be detected).

Subsequently, the trained models were used to examine data from the test sets. The accuracy for the entire set, the sensitivity (proportion of samples classified as correct positive = T2D), and specificity (proportion of samples classified as correct negative = healthy) were calculated. In Table 4, the accuracy for the validation and test sets of all prepared models described in Table 2, as well as the sensitivity and specificity for the corresponding test sets, are represented. All values are averaged with the results of the MCCV with three repetitions.

Table 4. Accuracy for the validation and test sets of all models at the genus level.

Model	Validation Set	Test Set	Specificity	Sensitivity
original	0.97 ± 0.00	0.96 ± 0.01	0.98 ± 0.01	0.94 ± 0.03
90°	0.95 ± 0.01	0.95 ± 0.01	0.96 ± 0.01	0.91 ± 0.04
180°	0.96 ± 0.01	0.96 ± 0.01	0.98 ± 0.02	0.91 ± 0.02
270°	0.96 ± 0.01	0.96 ± 0.00	0.98 ± 0.01	0.91 ± 0.02
vertical	0.96 ± 0.00	0.97 ± 0.01	0.98 ± 0.01	0.93 ± 0.05
horizontal	0.96 ± 0.01	0.96 ± 0.01	0.99 ± 0.01	0.92 ± 0.03
shuffled_a	0.94 ± 0.01	0.94 ± 0.01	0.95 ± 0.02	0.91 ± 0.03
shuffled_b	0.95 ± 0.01	0.95 ± 0.01	0.97 ± 0.01	0.89 ± 0.03
Ø	0.96 ± 0.01	0.96 ± 0.01	0.97± 0.01	0.92 ± 0.02

All accuracy values are very homogeneous between the different models and the standard deviations show only small fluctuations. Averaging all scores shows very good classification results for the validation and test sets of 0.96, as well as an outstanding specificity for the detection of the healthy class of 0.97 and a good result for the T2D class of 0.92 sensitivity for the test sets. To compare the results for all models, the following diagram (Figure 4) shows accuracy, specificity, and sensitivity values noted down in Table 4.

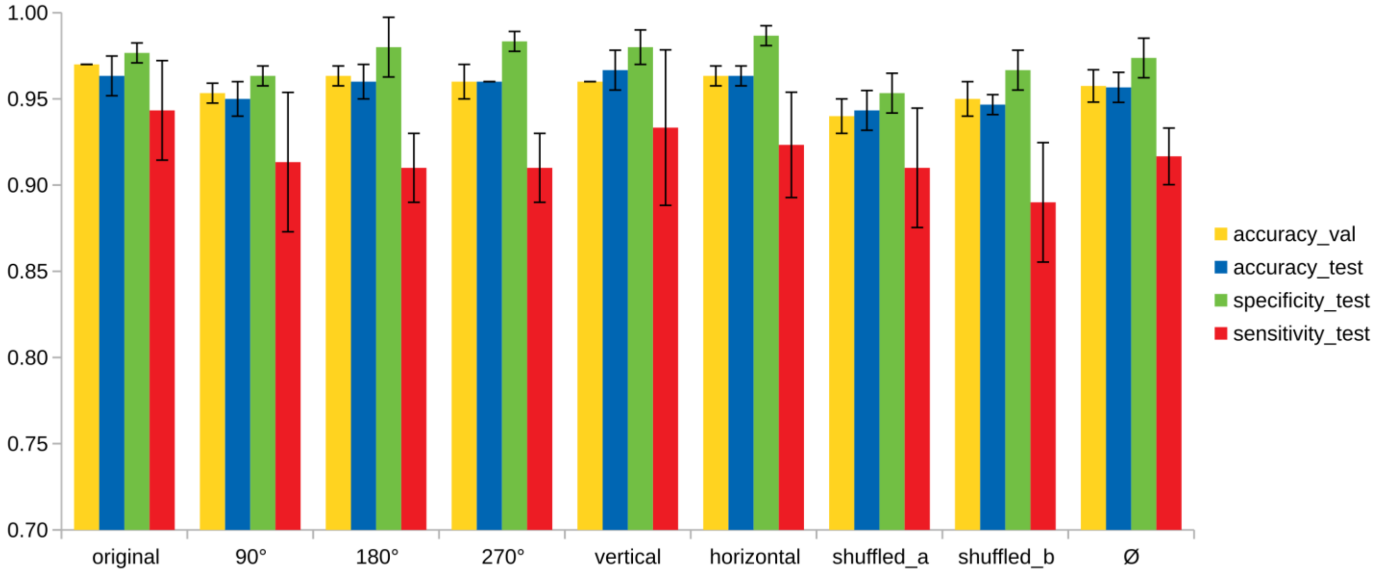

Figure 4. Bar chart of accuracy, specificity, and sensitivity values for the validation and test set. Results are presented for genus level data. For all datasets (original, 90°, 180°, 270°, vertical, horizontal, shuffled_a, and shuffled_b; details in Table 2) the accuracy for the validation and test set, and for the latter the associated 'healthy' and 'T2D' classes, are shown.

The average accuracy for the validation sets ranges between 0.94 and 0.97. For the test sets, the accuracy is slightly lower for most models because the neural networks were optimized using the validation sets. The specificity is described by values from 0.95 and 0.99 for all models. The sensitivity is somewhat lower at values between 0.89 and 0.94. The underlying reason for this is the smaller dataset for T2D (just under 30%) and possible other influencing factors, such as other previous diseases and severity of pathology. However, it is striking that the results are very homogeneous for all forms of visualization and arrangements of the genera. The MCCV with three repetitions also shows only minor fluctuations of a maximum of 0.05 in terms of standard deviation, so that a high robustness of the visualization and the models could be demonstrated. In Figure 5, an example confusion matrix underlines the performance on the test set from the original dataset. It is also recognizable that the small values for FP (0.02) and FN (0.06) demonstrate the classification accuracy.

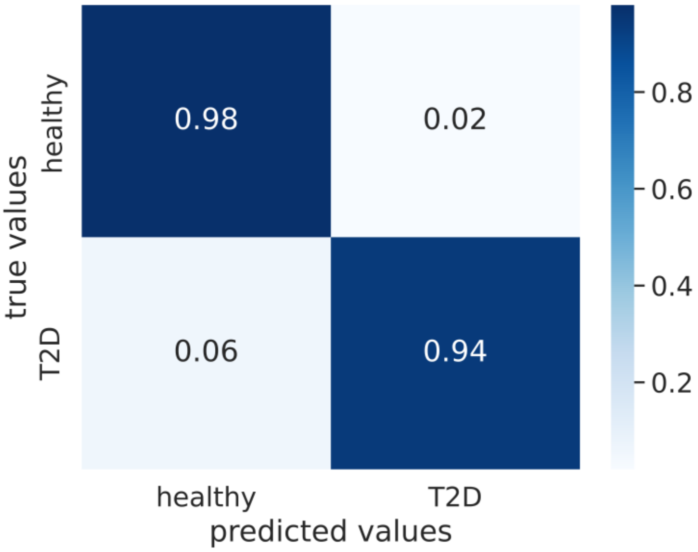

Figure 5. Confusion matrix of the test set from the original dataset.

The results for the mislabeled dataset do not show any significant classification results. Either all samples are classified as healthy or as T2D, resulting in an accuracy of 0.71 or 0.29, which represents the proportion of the whole dataset. This approach ensured that there were no other non-obvious factors influencing classification performance. The hypothesis that a random arrangement of intensities is responsible for the classification, and not the intensities themselves, could therefore be refuted.

The conventional ML model RF achieved an accuracy of 0.94, a specificity of 0.99, and a sensitivity of 0.76 for the disease detection. The SVM reached an accuracy of 0.86, a specificity of 0.89, and a sensitivity of 0.78. The image-classification-based models performed better in terms of overall accuracy and were superior in terms of sensitivity. Only in specificity did RF perform slightly better, but under the impression of a weak classification performance for the T2D group.

4. Discussion and Conclusions

The approach of visualizing microbiome sequencing data as radial heatmaps and analyzing them with the powerful DL-based ResNet-50 image classifier was proven to be beneficial for the utilized healthy/T2D dataset. This involved adapting the hyperparameters and training from scratch, as the network was optimized for lower resolution images and larger batch sizes [32–34].

Different visualizations at the genus level were used for training and classification to check robustness and generalization potential. The applied cross-validation and the comparison between validation and test set revealed no particularly advantageous visualization. The number of features seems to be of crucial importance in image classification. At the genus level (362 features), excellent results were achieved, in contrast to the outcome for the phylum (62 attributes) and class (115 attributes) levels, where the number of features seemed insufficient. The aggregation of different genera to higher taxonomic levels seems to result in a loss of important information for the classification, and individual genera in particular might be causative. The research work of Thambawita et al. [50] proved that an increased image resolution (which could be compared with an increased number of features) improves classification accuracy of DL models. Consequently, the orientation of the visualization and the order of the genera are irrelevant; only a sufficiently large number of the intensities must be represented as features.

Other scientific studies using CNNs for the classification of T2D achieved an accuracy of 0.68 [28] or maximum AUC values of 0.81 [27], 0.67 [29], and 0.75 [17]. Direct comparability is therefore not possible, but it can be assumed that for an AUC of 0.81, maximum sensitivity and specificity of 0.9 was reached. Thus, our approach achieved better results for the classification of T2D with a specificity of 0.97, a sensitivity of 0.92, and an accuracy of 0.96. Such a performance has not been achieved using ML methods such as RF and SVM, or with techniques from other research studies. Thus, the presented approach provides a simple and accurate classification. Problems with custom CNNs [25–29] include shallow depth architectures with few layers or classifier overfitting. ResNets did not achieve success because ImageNet [32] pretrained weights were used [27], or the data visualization as a bar chart of microbiome abundances [41] was simply not suitable for the network.

To evaluate our approach for microbiome image classification in more detail, datasets with other phenotypes (sampling site) and diseases (LC, CRC, obesity) should be investigated to assess the generalization potential to detect changes in microbial composition. Multi-class detection and the prediction of diseases are also conceivable. To reveal the causative genera for disease detection or their responsibility for being healthy, explainable artificial intelligence (XAI) approaches can be used. Methods such as class activation mapping (CAM) [51], local interpretable model-agnostic explanations (LIME) [52], or Shapley additive explanations (SHAP) [53] could clarify which features are responsible and help to obtain detailed insights into the pathogenesis. This could lead to the development of diagnostic tests for the detection and prediction of diseases, as well as instructions to prevent or retard their progression through directed adaptation of the microbiome.

Author Contributions: Conceptualization and methodology, J.P., J.S. and A.N.; software, visualization and formal analysis, J.P. and J.S.; validation and investigation, J.P.; resources and data curation, O.M. and M.Y.; writing—original draft preparation, J.P., J.S. and A.N.; writing—review and editing, M.F.; supervision, H.P., F.T.H. and M.F.; project administration, M.F.; funding acquisition, J.P. and M.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research and the APC was funded by the Federal Ministry of Education and Research (BMBF) of Germany under grant number 13FH209PX8.

Institutional Review Board Statement: The data are provided by the company BIOMES NGS GmbH. Customers pay for the microbiome test and give their informed consent for further scientific use in the purchase process. All data have been anonymized. The study and all methods were performed in accordance with the Declaration of Helsinki and in line with §32, which states that an appropriate ethics approval is only required for identifiable human material or data. The statement of the German central ethics committee (panel of the German medical association) indicates that in this context in accordance with the Declaration of Helsinki, ethics approval is not considered necessary for anonymized human material/data [54] (p. 16).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the fact that the research project was conducted in collaboration with the company BIOMES NGS GmbH. The collaborating company has a vested interest in the success of the research in the field of microbiome analysis, therefore it is not possible that the data and code are fully publicly available.

Acknowledgments: Special thanks to the laboratory staff of BIOMES NGS GmbH. Many thanks also to all reviewers for their corrections, comments, and endeavors.

Conflicts of Interest: The authors declare no conflict of interest. The submitted work was performed in cooperation with the company BIOMES NGS GmbH, Wildau, Germany. They work in the field of microbiome analysis and have therefore a vested interest in the success of this field. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Fan, Y.; Pedersen, O. Gut Microbiota in Human Metabolic Health and Disease. *Nat. Rev. Microbiol.* **2021**, *19*, 55–71. [CrossRef]
2. Akimbekov, N.S.; Digel, I.; Sherelkhan, D.K.; Lutfor, A.B.; Razzaque, M.S. Vitamin D and the Host-Gut Microbiome: A Brief Overview. *Acta Histochem. Cytochem.* **2020**, *53*, 33–42. [CrossRef] [PubMed]
3. Davis, C.D. The Gut Microbiome and Its Role in Obesity. *Nutr. Today* **2016**, *51*, 167. [CrossRef] [PubMed]
4. Gevers, D.; Kugathasan, S.; Denson, L.A.; Vázquez-Baeza, Y.; Van Treuren, W.; Ren, B.; Schwager, E.; Knights, D.; Song, S.J.; Yassour, M. The Treatment-Naive Microbiome in New-Onset Crohn's Disease. *Cell Host Microbe* **2014**, *15*, 382–392. [CrossRef] [PubMed]
5. Sears, C.L.; Garrett, W.S. Microbes, Microbiota, and Colon Cancer. *Cell Host Microbe* **2014**, *15*, 317–328. [CrossRef]
6. Xu, H.; Liu, M.; Cao, J.; Li, X.; Fan, D.; Xia, Y.; Lu, X.; Li, J.; Ju, D.; Zhao, H. The Dynamic Interplay between the Gut Microbiota and Autoimmune Diseases. *J. Immunol. Res.* **2019**, *2019*, 351–364. [CrossRef]
7. Zhang, H.; Chen, Y.; Wang, Z.; Xie, G.; Liu, M.; Yuan, B.; Chai, H.; Wang, W.; Cheng, P. Implications of Gut Microbiota in Neurodegenerative Diseases. *Front. Immunol.* **2022**, *13*, 325. [CrossRef] [PubMed]
8. Hasan, N.; Yang, H. Factors Affecting the Composition of the Gut Microbiota, and Its Modulation. *PeerJ* **2019**, *7*, e7502. [CrossRef] [PubMed]
9. Loncar-Turukalo, T.; Claesson, M.J.; Bertelsen, R.J.; Zomer, A.; D'Elia, D. Towards the Optimisation and Standardisation of Machine Learning Techniques for Human Microbiome Research: The ML4Microbiome COST Action (CA 18131). *EMBnet J.* **2020**, *26*, 997. [CrossRef]
10. Bansal, N. Prediabetes Diagnosis and Treatment: A Review. *World J. Diabetes* **2015**, *6*, 296. [CrossRef]
11. Wukich, D.K.; Raspovic, K.M.; Suder, N.C. Patients with Diabetic Foot Disease Fear Major Lower-Extremity Amputation More than Death. *Foot Ankle Spec.* **2018**, *11*, 17–21. [CrossRef]
12. Wensel, C.R.; Pluznick, J.L.; Salzberg, S.L.; Sears, C.L. Next-Generation Sequencing: Insights to Advance Clinical Investigations of the Microbiome. *J. Clin. Investig.* **2022**, *132*, e154944. [CrossRef]
13. Douglas, G.M.; Maffei, V.J.; Zaneveld, J.R.; Yurgel, S.N.; Brown, J.R.; Taylor, C.M.; Huttenhower, C.; Langille, M.G. PICRUSt2 for Prediction of Metagenome Functions. *Nat. Biotechnol.* **2020**, *38*, 685–688. [CrossRef]
14. Moreno-Indias, I.; Lahti, L.; Nedyalkova, M.; Elbere, I.; Roshchupkin, G.; Adilovic, M.; Aydemir, O.; Bakir-Gungor, B.; Santa Pau, E.C.; D'Elia, D. Statistical and Machine Learning Techniques in Human Microbiome Studies: Contemporary Challenges and Solutions. *Front. Microbiol.* **2021**, *12*, 277. [CrossRef]
15. Shwartz-Ziv, R.; Armon, A. Tabular Data: Deep Learning Is Not All You Need. *Inf. Fusion* **2022**, *81*, 84–90. [CrossRef]
16. Namkung, J. Machine Learning Methods for Microbiome Studies. *J. Microbiol.* **2020**, *58*, 206–216. [CrossRef]
17. Sharma, D.; Paterson, A.D.; Xu, W. TaxoNN: Ensemble of Neural Networks on Stratified Microbiome Data for Disease Prediction. *Bioinformatics* **2020**, *36*, 4544–4550. [CrossRef]
18. Reitmeier, S.; Kiessling, S.; Clavel, T.; List, M.; Almeida, E.L.; Ghosh, T.S.; Neuhaus, K.; Grallert, H.; Linseisen, J.; Skurk, T. Arrhythmic Gut Microbiome Signatures Predict Risk of Type 2 Diabetes. *Cell Host Microbe* **2020**, *28*, 258–272. [CrossRef]
19. Hernández Medina, R.; Kutuzova, S.; Nielsen, K.N.; Johansen, J.; Hansen, L.H.; Nielsen, M.; Rasmussen, S. Machine Learning and Deep Learning Applications in Microbiome Research. *ISME Commun.* **2022**, *2*, 98. [CrossRef]
20. Mulenga, M.; Kareem, S.A.; Sabri, A.Q.M.; Seera, M.; Govind, S.; Samudi, C.; Mohamad, S.B. Feature Extension of Gut Microbiome Data for Deep Neural Network-Based Colorectal Cancer Classification. *IEEE Access* **2021**, *9*, 23565–23578. [CrossRef]
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
22. Sarwinda, D.; Paradisa, R.H.; Bustamam, A.; Anggia, P. Deep Learning in Image Classification Using Residual Network (ResNet) Variants for Detection of Colorectal Cancer. *Procedia Comput. Sci.* **2021**, *179*, 423–431. [CrossRef]
23. O'Mahony, N.; Campbell, S.; Carvalho, A.; Harapanahalli, S.; Hernandez, G.V.; Krpalkova, L.; Riordan, D.; Walsh, J. Deep Learning vs. Traditional Computer Vision. In *Advances in Computer Vision, Proceedings of the 2019 Computer Vision Conference (CVC), Las Vegas, NV, USA, 2–3 May 2019*; Springer: Cham, Switzerland, 2019; pp. 128–144.
24. Wang, W.; Yang, Y.; Wang, X.; Wang, W.; Li, J. Development of Convolutional Neural Network and Its Application in Image Classification: A Survey. *Opt. Eng.* **2019**, *58*, 040901. [CrossRef]
25. Reiman, D.; Metwally, A.; Dai, Y. Using Convolutional Neural Networks to Explore the Microbiome. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, Republic of Korea, 11–15 July 2017; IEEE: New York, NY, USA, 2017; pp. 4269–4272.
26. Reiman, D.; Farhat, A.M.; Dai, Y. Predicting Host Phenotype Based on Gut Microbiome Using a Convolutional Neural Network Approach. In *Artificial Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 249–266. [CrossRef]
27. Chen, X.; Zhu, Z.; Zhang, W.; Wang, Y.; Wang, F.; Yang, J.; Wong, K.-C. Human Disease Prediction from Microbiome Data by Multiple Feature Fusion and Deep Learning. *iScience* **2022**, *25*, 104081. [CrossRef]
28. Nguyen, T.H.; Prifti, E.; Sokolovska, N.; Zucker, J.-D. Disease Prediction Using Synthetic Image Representations of Metagenomic Data and Convolutional Neural Networks. In Proceedings of the 2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF), Danang, Vietnam, 20–22 March 2019; IEEE: New York, NY, USA, 2019; pp. 1–6.

29. Li, B.; Zhong, D.; Jiang, X.; He, T. TopoPhy-CNN: Integrating Topological Information of Phylogenetic Tree for Host Phenotype Prediction from Metagenomic Data. In Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Houston, TX, USA, 9–12 December 2021; IEEE: New York, NY, USA, 2021; pp. 456–461.
30. Sharma, A.; Vans, E.; Shigemizu, D.; Boroevich, K.A.; Tsunoda, T. DeepInsight: A Methodology to Transform a Non-Image Data to an Image for Convolution Neural Network Architecture. *Sci. Rep.* **2019**, *9*, 11399. [CrossRef]
31. Bruno, P.; Calimeri, F. Using Heatmaps for Deep Learning Based Disease Classification. In Proceedings of the 2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Siena, Italy, 9–11 July 2019; IEEE: New York, NY, USA, 2019; pp. 1–7.
32. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. Imagenet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; IEEE: New York, NY, USA, 2009; pp. 248–255.
33. Deng, L. The Mnist Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]. *IEEE Signal Process. Mag.* **2012**, *29*, 141–142. [CrossRef]
34. Krizhevsky, A.; Hinton, G. Learning Multiple Layers of Features from Tiny Images. 2009. Available online: http://www.cs.utoronto.ca/~kriz/learning-features-2009-TR.pdf (accessed on 14 February 2023).
35. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
36. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
37. Lu, Z.; Lu, J.; Ge, Q.; Zhan, T. Multi-Object Detection Method Based on YOLO and ResNet Hybrid Networks. In Proceedings of the 2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM), Toyonaka, Japan, 3–5 July 2019; IEEE: New York, NY, USA, 2019; pp. 827–832.
38. Sanchez, S.; Romero, H.; Morales, A. A Review: Comparison of Performance Metrics of Pretrained Models for Object Detection Using the TensorFlow Framework. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *844*, 012024. [CrossRef]
39. Lu, Y.; Qin, X.; Fan, H.; Lai, T.; Li, Z. WBC-Net: A White Blood Cell Segmentation Network Based on UNet++ and ResNet. *Appl. Soft Comput.* **2021**, *101*, 107006. [CrossRef]
40. Pfeil, J.; Nechyporenko, A.; Frohme, M.; Hufert, F.T.; Schulze, K. Examination of Blood Samples Using Deep Learning and Mobile Microscopy. *BMC Bioinform.* **2022**, *23*, 65. [CrossRef]
41. Michel-Mata, S.; Wang, X.; Liu, Y.; Angulo, M.T. Predicting Microbiome Compositions from Species Assemblages through Deep Learning. *iMeta* **2022**, *1*, e3. [CrossRef]
42. Siptroth, J.; Moskalenko, O.; Krumbiegel, C.; Ackermann, J.; Koch, I.; Pospisil, H. Variation of Butyrate Production in the Gut Microbiome in Type 2 Diabetes Patients. *Int. Microbiol.* **2023**. [CrossRef] [PubMed]
43. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data. *Bioinformatics* **2012**, *28*, 3150–3152. [CrossRef] [PubMed]
44. Li, W.; Godzik, A. Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [CrossRef] [PubMed]
45. Waskom, M.L. Seaborn: Statistical Data Visualization. *J. Open Source Softw.* **2021**, *6*, 3021. [CrossRef]
46. Arlot, S.; Celisse, A. A Survey of Cross-Validation Procedures for Model Selection. *Stat. Surv.* **2010**, *4*, 40–79. [CrossRef]
47. Chollet, F. Keras: The Python Deep Learning Library. *Astrophys. Source Code Libr.* **2018**, *2018*, ascl-1806.
48. Liaw, A.; Wiener, M. Classification and Regression by RandomForest. *R News* **2002**, *2*, 18–22.
49. Suykens, J.A.; Vandewalle, J. Least Squares Support Vector Machine Classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300. [CrossRef]
50. Thambawita, V.; Strümke, I.; Hicks, S.A.; Halvorsen, P.; Parasa, S.; Riegler, M.A. Impact of Image Resolution on Deep Learning Performance in Endoscopy Image Classification: An Experimental Study Using a Large Dataset of Endoscopic Images. *Diagnostics* **2021**, *11*, 2183. [CrossRef]
51. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning Deep Features for Discriminative Localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.
52. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why Should i Trust You?" Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144.
53. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 7874. [CrossRef]
54. Zentrale Ethikkommission Stellungnahme Der Zentralen Ethikkommission. Die (Weiter-)Verwendung von Menschlichen Körpermaterialien Für Zwecke Medizinischer Forschung (20.02.2003). Available online: https://www.zentrale-ethikkommission.de/fileadmin/user_upload/_old-files/downloads/pdf-Ordner/Zeko/Koerpermat-1.pdf (accessed on 22 December 2022).

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.