applied sciences

MDPI

Article

Innovative Approaches to Clinical Diagnosis: Transfer Learning in Facial Image Classification for Celiac Disease Identification

Elif Keskin Bilgiç [1], İnci Zaim Gökbay [2,3,*] and Yusuf Kayar [4]

[1] Department of Biomedical Engineering, Faculty of Engineering, Istanbul Cerrahpaşa University, 34320 Istanbul, Turkey; elif.keskin@ogr.iuc.edu.tr
[2] Informatics Department, İstanbul University, 34134 Istanbul, Turkey
[3] Institute of Cyber Security for Society (iCSS) & School of Computing, University of Kent, Canterbury CT2 7NP, UK
[4] Gastroenterology Department, Van Education and Research Hospital, 65300 Van, Turkey; ykayar@yahoo.com
* Correspondence: inci.gokbay@istanbul.edu.tr

Abstract: Background: Celiac disease arises from gluten consumption and shares symptoms with other conditions, leading to delayed diagnoses. Untreated celiac disease heightens the risk of autoimmune disorders, neurological issues, and certain cancers like lymphoma while also impacting skin health due to intestinal disruptions. This study uses facial photos to distinguish individuals with celiac disease from those without. Surprisingly, there is a lack of research involving transfer learning for this purpose despite its benefits such as faster training, enhanced performance, and reduced overfitting. While numerous studies exist on endoscopic intestinal photo classification and a few have explored the link between facial morphology measurements and celiac disease, none have concentrated on diagnosing celiac disease through facial photo classification. Methods: This study sought to apply transfer learning techniques with VGG16 to address a gap in research by identifying distinct facial features that differentiate patients with celiac disease from healthy individuals. A dataset containing a total of 200 facial images of adult individuals with and without celiac condition was utilized. Half of the dataset had a ratio of 70% females to 30% males with celiac condition, and the rest had a ratio of 60% females to 40% males without celiac condition. Among those with celiac condition, 28 were newly diagnosed and 72 had been previously diagnosed, with 25 not adhering to a gluten-free diet and 47 partially adhering to such a diet. Results: Utilizing transfer learning, the model achieved a 73% accuracy in classifying the facial images of the patients during testing, with corresponding precision, recall, and F1 score values of 0.54, 0.56, and 0.52, respectively. The training process involved 50,178 parameters, showcasing the model's efficacy in diagnostic image analysis. Conclusions: The model correctly classified approximately three-quarters of the test images. While this is a reasonable level of accuracy, it also suggests that there is room for improvement as the dataset contains images that are inherently difficult to classify even for humans. Increasing the proportion of newly diagnosed patients in the dataset and expanding the dataset size could notably improve the model's efficacy. Despite being the first study in this field, further refinement holds promise for the development of a diagnostic tool for celiac disease using transfer learning in medical image analysis, addressing the lack of prior studies in this area.

Keywords: celiac disease diagnosis; image classification; deep learning; transfer learning; VGG16 pre-trained model

check for updates

Citation: Keskin Bilgiç, E.; Zaim Gökbay, İ.; Kayar, Y. Innovative Approaches to Clinical Diagnosis: Transfer Learning in Facial Image Classification for Celiac Disease Identification. *Appl. Sci.* **2024**, *14*, 6207. https://doi.org/10.3390/app14146207

Academic Editor: Maria Rizzi

Received: 29 April 2024
Revised: 8 July 2024
Accepted: 9 July 2024
Published: 17 July 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

Celiac disease (CD) is an autoimmune disorder primarily affecting the small intestine, triggered by the ingestion of gluten—a protein found in wheat, barley, and rye [1]. Upon gluten consumption, individuals with CD experience an immune response that damages the small intestine's lining, leading to various symptoms and potential long-term complications [2]. A definitive diagnosis often necessitates a small intestine biopsy to assess the

extent of damage. However, it typically takes around 4–5 years for physicians to decide to recommend a biopsy. During this time, symptoms worsen and become more challenging [3]. Addressing these limitations requires ongoing research to improve the accuracy and timeliness of CD diagnosis. CD diagnosis is challenging due to its diverse clinical manifestations and symptoms that are shared with other diseases [4]. Delayed diagnosis can exacerbate disease progression and intestinal permeability in its early stages [5].

In the present era, there is a heightened emphasis on discovering automated computational mechanisms that can aid in [6] the determination of intestinal atrophy in celiac disease, but there is a lot less research on the impact of gluten sensitivity on facial appearance. CD is associated with impairments in gut health that can extend to the skin through a variety of mechanisms, including inflammation, immune system changes, and microbiome dysbiosis [7,8]. The craniofacial morphology of patients with celiac disease has been shown to reveal a distinct pattern of craniofacial growth [9]. Moreover, celiac patients often exhibit distinctive facial features, such as paleness due to anemia, itchy rashes with small blisters around the hairline and mouth (dermatitis herpetiformis), and inflammation and cracking at the corners of the lips (cheilitis) [10]. There may also be tooth discoloration, tooth enamel defects such as pitting or cupping, facial edema (swelling around the eyes), dark circles under the eyes due to chronic fatigue or a lack of nutrients, and a gaunt, hollow appearance with reduced facial fat due to weight loss [11]. The facial features described above are the features we expect the algorithms to detect in our study.

Facial image classification holds promise in contributing to the diagnosis and identification of specific diseases, especially those with observable facial features or patterns [12]. Given that CD primarily affects the gastrointestinal tract and can result in nutrient malabsorption, it may manifest in changes in skin and facial characteristics [13].

In the medical imaging field, acquiring a sufficiently large dataset to train a deep convolutional neural network (CNN) from scratch poses a significant challenge. Training a network from scratch requires substantial computational and memory resources, along with extensive datasets, which are often unavailable. However, when dealing with small datasets, the risk of encountering overfitting issues becomes more prevalent [14]. To address this challenge, transfer learning emerges as a viable solution.

This paper adopts a transfer learning approach, specifically utilizing VGG16, as a tool for customizing existing machine learning (ML) models to work with new datasets [15]. VGG16 has been used as a pretrained model to fine-tune CD face image classifications. The VGG16 model, known for its efficiency in mobile vision applications, is employed for fine-tuning image datasets [16]. Subsequently, it can serve as an inference engine, providing the flexibility to infer required information from new CD images when necessary [17]. The study included volunteers over the age of 18 diagnosed with celiac disease. The research was conducted through one-on-one interviews and used the participants' photographs.

A dataset comprising 100 images of patients diagnosed with celiac disease (CD) was assembled by the Gastroenterology department of Van Education and Research Hospital over the period of 2022–2023. The dataset was meticulously curated following protocols established on 2022/09-02 to ensure data integrity and quality. To complement the CD patient images, an additional set of 100 images depicting non-celiac patients was procured from the same hospital. This balanced approach in dataset collection enabled a comprehensive analysis of facial characteristics relevant to celiac disease diagnosis. The research endeavor was supported in part by the Health Science University in Turkey, fostering collaboration and advancements in medical research. This study represents original research, pioneering new paths within its field. The researchers conducted their own experiments, gathering fresh data to address previously unanswered questions. These efforts have yielded novel insights, paving the way for further investigation.

2. Background and Related Work

2.1. Machine Learning and Deep Neural Network

Machine learning, a field in which algorithms facilitate predictions, hinges on a machine's ability to learn through extensive data exposure [18].

Classification algorithms, including support vector machines, naïve Bayes, hidden Markov models, Bayesian networks, and neural networks, are pivotal in this domain [19].

In another category of classification algorithms, convolutional neural networks (CNNs) are discriminative, supervised models trained using backpropagation and gradient descent [20,21]. Recent advancements in CNN architecture prioritize efficient resource utilization, considering factors like memory, power budget, and execution speed. VGG16 is a convolutional neural network (CNN) architecture that was developed by the Visual Geometry Group (VGG) at the University of Oxford. It was introduced in the paper titled "Very Deep Convolutional Networks for Large-Scale Image Recognition" by Karen Simonyan and Andrew Zisserman in 2014 [22].

Research findings underscore VGG16's effectiveness as a feature extractor for image classification, object detection, and segmentation [23].

2.2. Transfer Learning

In the context of deep neural networks, it is uncommon to have a sufficiently large dataset to train an entire convolutional neural network (CNN) from scratch. Deep neural networks, with their numerous parameters, face challenges when trained on small datasets, impacting their ability to generalize and often resulting in overfitting [18].

Transfer learning (TL) originates from cognitive research, wherein the concept revolves around the notion that knowledge can be transferred between related tasks to enhance performance in a novel task. It is widely acknowledged that humans excel at solving similar tasks by leveraging their previous knowledge [24].

Decision support systems leveraging deep learning (DL), machine learning (ML), and neural networks (NNs) exhibit a high level of efficiency but are hindered in the medical field by limited comprehensive datasets. Despite advancements in machine learning, the development of DSSs for CD diagnosis is relatively unexplored, likely due to the complexity of the issue and the scarcity of public databases.

Gadermayr et al. [25] provide an overview of contemporary developments in computer-aided diagnoses of celiac disease utilizing upper endoscopies. They suggest frameworks for achieving fully automated patient-specific diagnoses and for incorporating expert knowledge into the automated decision-making process.

In [26], the authors presented a feature descriptor tailored for the classification of video capsule endoscopy images. Furthermore, they introduced a system that utilizes deep convolutional neural networks (CNNs) to characterize small intestine motility by detecting individual motility events.

In recent years, deep learning methods, including prominent convolutional neural networks like AlexNet, GoogLeNet, and VGG16 net, have been increasingly utilized for the classification of endoscopic images.

Wang et al. [27] propose a deep-learning-based approach aimed at recalibrating modules to distinguish images featuring regions significant for celiac disease from images of healthy ones. Their method, employing support vector machine and k-nearest neighbor modules for a linear discriminant analysis, achieved impressive results of a 95.94% accuracy, 97.20% sensitivity, and 95.63% specificity using a 10-time 10-fold cross-validation strategy.

Amirkhani et al. [28] devised a methodology utilizing a fuzzy cognitive map (FCM) alongside a possibilistic fuzzy C-means clustering algorithm (PFCM) to categorize celiac disease (CD). Their research objective aimed at developing an expert system capable of classifying CD patients into three grades: A, B1, and B2, representing the latest available grading method. Three experts identified seven key defining features of CD, which were treated as FCM concepts. The authors achieved accuracy rates of 88%, 90%, and 91% for

the respective analysis classes. The experimental results revealed an impressive mean classification accuracy of 96% for six distinct intestinal motility events.

To our understanding, there has been a lack of development in clinical support systems tailored specifically for celiac disease screening, particularly those intended to operate based on non-invasive diagnostic tests with the aim of reducing reliance on biopsies. This gap in research likely stems from both the scarcity of screening data and the intricate nature of the diagnostic challenge. This study aims to investigate whether transfer learning algorithms can achieve a sufficient accuracy in classifying facial images to aid primary care physicians in making referrals to specialists for the diagnosis of CD. VGG16 architecture was employed in transfer learning for classifying facial images of people with celiac disease to aid the diagnosis of CD.

VGG16 is a pre-trained model known for its simplicity and efficiency in learning hierarchical features from images through its deep convolutional layers [29].

Transfer learning allows the VGG16 model, which is trained on large datasets such as ImageNet, to be efficiently fine-tuned with new data, reducing the training time and computational resources (Figure 1) [30].

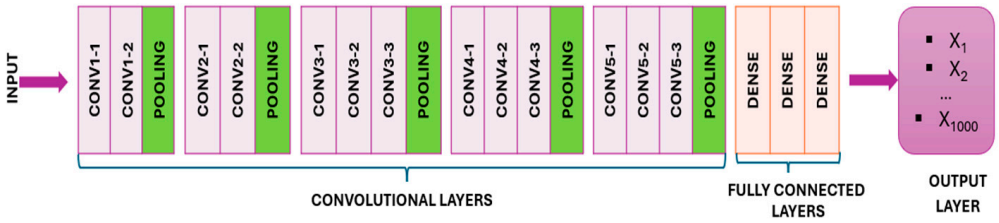


Figure 1. The VGG16 model.

VGG16 has a depth of 16 layers consisting of 13 convolutional layers and three fully connected layers with 3 × 3 filters and max pooling for dimensionality reduction and robustness. The network ends with fully connected layers and a SoftMax layer for classification. Despite its effectiveness, VGG16's 138 million parameters pose a computational challenge [31]. Through the technique of transfer learning, VGG16 capitalizes on the foundational logic already embedded within pre-existing models, expediting the training process and requiring fewer data for effective retraining [17].

In this approach, the architecture of a pre-trained model is utilized to generate a new dataset from input images. Specifically, the convolutional and pooling layers of the model are imported, excluding the "top portion" containing the fully connected layer. By passing images through these layers of the VGG16 model, a feature stack representing detected visual features is obtained. Subsequently, this three-dimensional feature stack is converted into a NumPy array, facilitating its seamless integration with various modeling techniques (Figure 2) [32,33].

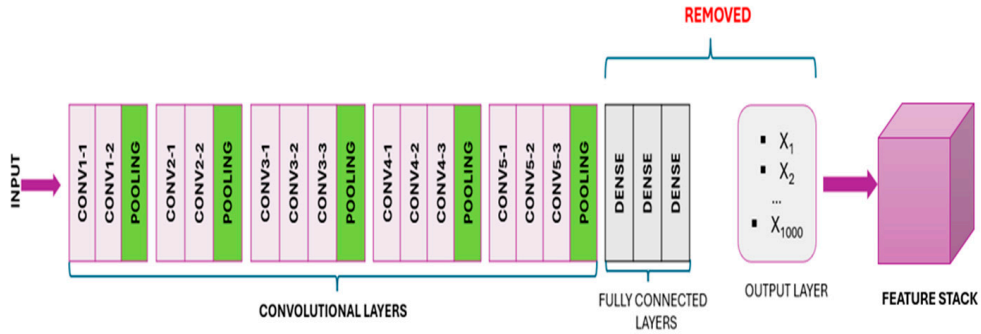


Figure 2. After removing the "top" portion, the model convolves the pre-trained output layers, resulting in a 3D stack of feature maps.

The pre-trained layers of VGG16 were utilized for feature extraction, in which the image dataset underwent processing through the convolutional layers and weights to generate transformed visual features. No direct training was performed on these pre-trained layers. While the frozen pre-trained layers maintained their usual convolution of visual features, the non-frozen (or 'trainable') pre-trained layers were adapted to our custom dataset, updating in accordance with predictions from the fully connected layer (Figure 3) [30].

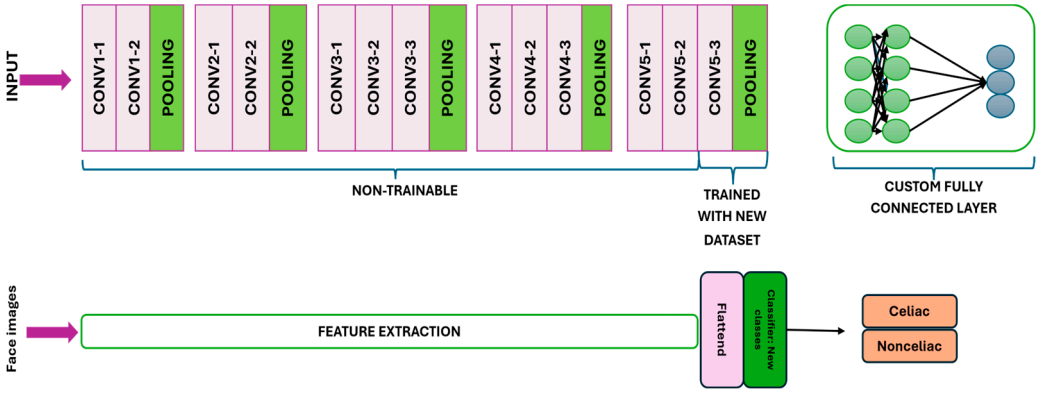


Figure 3. The final convolutional and pooling layers are unfrozen to enable training, while a fully connected layer is established for both training and prediction purposes.

In a pre-trained model like VGG16, we only train additional classifier layers, which is advantageous for several reasons as follows.

Transfer learning efficiency: It uses pre-trained weights to extract general features and increases transfer learning efficiency. Overfitting prevention: Freezing pre-trained layers provides better generalizations and reduces the risk of overfitting on smaller datasets. Training speed: It reduces the overall training time and computational costs by focusing computational resources on new layers. Therefore, instead of retraining the entire model, we only tune additional classifier layers, as this is more time-efficient and requires fewer computational resources.

The VGG16 model was selected due to its proven track record in image classification tasks and its deep architecture which enables it to capture complex features in images. Using VGG16 pre-trained weights on the ImageNet dataset provides a solid foundation that allows transfer learning to adapt to specific tasks with a smaller dataset. This significantly reduces the training time and required computational resources while also increasing the model's ability to generalize from a relatively small dataset by leveraging the features learned from a large and diverse set of images. Resnet was also used in our preliminary work. Resnet is a pre-trained transfer learning model like VGG16. VGG16 was preferred due to the low performance of Resnet.

3. Materials and Method

Research Design: The last layer of the VGG16 pre-trained model was trained only with the celiac patient dataset and made a prediction by classifying the newly entered photo (Figure 4). The weights of VGG16 were not changed.

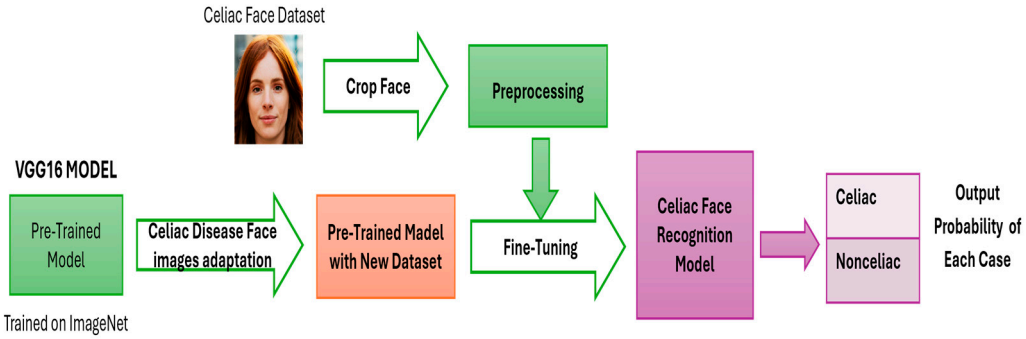


Figure 4. Proposed work.

3.1. Collection and Preparation of Datasets

The dataset encompasses two categories for the final model to recognize. Compared to training a model from scratch, the dataset required for transfer learning can be significantly smaller. Transfer learning with VGG16 enables the retraining of an existing model with a reduced dataset while still achieving satisfactory performance. In these instructions, we utilized a smaller dataset featuring 2 classes, each containing 100 images. A collection of CD and NCD image datasets was obtained from Van Training and Research Hospital, Department of Gastroenterology (Figure 5). When a gluten-free diet is followed, the manifestations mentioned above in the introduction decrease or disappear completely, as can be seen in clinical tests. For this reason, newly diagnosed and previously diagnosed patients who were not on a diet with gluten or who were on a partial gluten-free diet were included in the study. Information about the diet status of celiac patients is shown in Figure 6.

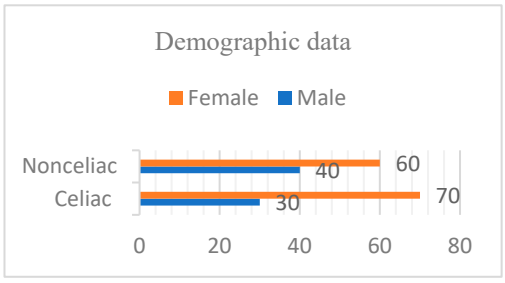


Figure 5. Males and females in both groups.

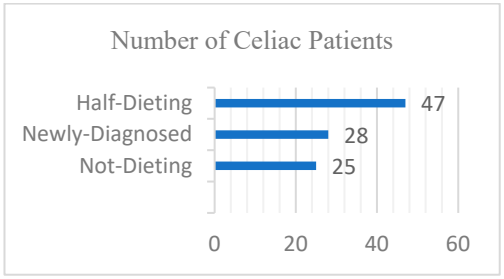


Figure 6. Diet status of celiac patients.

Patients were informed by the specialist physician in accordance with the ethics committee and had to sign a form to ensure their participation was voluntary. The participants were positioned in front of a white wall while seated on a stationary chair (non-rotating standard-type). Using a Sony DSC-S2100 12.1MP camera (Tokyo, Japan) mounted on a tripod positioned 50 cm away, only vertical frontal images of their faces were captured, and we ensured that they were not wearing glasses or hats. Tripod: The lens height was at the

same level as the patients' faces, and the images taken covered their faces and stopped at the edges of their hair, shawls, etc. (Figure 7).

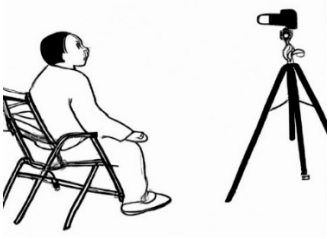


Figure 7. Position of patient and camera.

3.2. The Workflow

The model is fine-tuned on a custom dataset for image classification with the proposed work (Figure 4). The workflow includes image preprocessing, model configuration, adding custom layers, compilation, data augmentation, data generator, training loop, evaluation, visualization of results, and classification report. TensorFlow (v.2.16.1) and Keras (v.3.4.1) were used to create and train the deep learning model. Glob library was used to retrieve file paths from dataset directories. Matplotlib (v.3.5.3) and Seaborn (v.0.13.2) were used to plot training/validation metrics and confusion matrices. Scikit-learn was used to generate classification reports and confusion matrices.

3.2.1. Image Preprocessing

The original facial images were acquired with a size of at 1600 × 2000 pixels and were cropped to 400 × 600 pixels during data preparation step to focus solely on the faces. Pre-trained models for transfer learning resized the images to 224 × 224 pixels. While this resize is generally sufficient for many tasks involving facial features, it can introduce artifacts and obscure fine details. Pre-trained models like VGG16 help reduce detail loss by extracting robust features from low-resolution images. All images were resized to 224 × 224 pixels, which is the input size for VGG16. The paths for training and validation datasets are specified. The color channels of the images were reordered from RGB (red, green, blue) to BGR (blue, green, red). This is because VGG16 was trained on ImageNet using the BGR color format.

3.2.2. Model Configuration

The VGG16 model is loaded with pre-trained weights from the ImageNet dataset. To customize the model to a specific task, fully connected layers at the top are excluded (include_top = False).

Layer freezing: All pre-trained layers are frozen to prevent their weights from being updated during training, preserving the features learned from ImageNet.

3.2.3. Adding Custom Layers

To modify the VGG16 model for our classification task, we first flattened the 3D output of the VGG16 model into a 1D tensor using Flatten()(vgg.output). We then added two fully connected layers: a hidden layer with 1000 neurons and ReLU activation (Dense(1000, activation = 'relu')(x)) and an output layer with the number of neurons equal to the number of classes using softmax activation function for multiple operations (prediction (Dense(len(folders), activation = 'softmax')(x))). The modified model was defined with VGG16 input and custom output layers (Model(inputs = vgg.input, Outputs = prediction)) and summarized using model.summary().

3.2.4. Compilation

The model is compiled using categorical cross-entropy as the loss function, which is suitable for multi-class classification. Adam optimizer, known for its adaptive learning rate, is used to improve the training process. Additionally, accuracy is used as

a metric to evaluate the performance of the model. Compilation parameters are defined as follows: model.compile(loss = 'categorical_crossentropy', optimizer = 'adam', metrics = ['accuracy']).

3.2.5. Data Augmentation

The model was trained using a batch size of 32, processing 32 samples before updating the internal parameters. The training was conducted over 10 epochs, in which each epoch represents a complete pass through the entire training dataset. The learning rate, implicitly set by the Adam optimizer, was 0.001, which determines the step size at each iteration while moving towards the minimum of the loss function. The Adam optimizer was selected for updating the model parameters due to its efficiency and adaptive learning rate capabilities. The categorical cross entropy loss function was used to measure the discrepancy between the predicted labels and the true labels. To enhance the model's robustness and prevent overfitting, data augmentation techniques were employed, including rescaling pixel values by 1./255, applying random shearing transformations with a range of 0.2, random zooming within a range of 0.2, and horizontal flipping of images.

3.2.6. Data Generators

Efficient memory usage is facilitated by leveraging data generators to create images and tags in bulk directly from the directory. The training set was created using train_datagen.flow_from_directory specifying the training path, 224 × 224 target size, batch size, and categorical class mode. Similarly, the test set was created using test_datagen.flow_from_directory with validation path, target size of 224 × 224, batch size, and categorical class mode.

3.2.7. Training Loop

A custom training cycle was implemented to iterate over periods and groups, manually managing the training and validation processes for greater control and potential customizations. In each epoch, the model was trained in batches using model.train_on_batch(X_batch, y_batch) to obtain the training loss and accuracy. Validation loss and accuracy were evaluated using model.evaluate(test_set, steps = validation_steps) at the end of each epoch.

3.2.8. Evaluation and Metrics

The performance of the model was evaluated on the test set. Test loss and test results were calculated using the model.evaluate(test_set) function. The trained model is edited with the model.save('facefeatures_new_model.h5') file for future use.

3.2.9. Visualization

To visualize the training and validation process, graphs were created showing losses and accuracy by period. The first subplot shows the training and validation loss using plt.plot(epochs_range, training_losses, label = 'Training Loss') and plt.plot(epochs_range, validation_losses, label = 'Validation Loss'). The second subplot shows the training and validation accuracy with plt.plot(epochs_range, training_accuracies, label = 'Training Accuracy') and plt.plot(epochs_range, validation_accuracies, label = 'Validation Accuracy'). These plots were arranged side by side using plt.subplot and are displayed using plt.show().

3.2.10. Classification Report and Confusion Matrix

Predictions were made on the test set using Model.predict(test_set, steps = validation_steps), and the predicted classes were determined by np.argmax(Y_pred, axis = 1). A classification report was generated for each class using class_report(y_true, y_pred, target_names = target_names), elaboration precision, recall, and F1 score. To further evaluate the performance of the classifier in distinguishing classes, a confusion matrix was created using confusion_matrix(y_true, y_pred). This matrix was visualized with a

heatmap using sns.heatmap(conf_matrix, annot = True, fmt = 'd', xticklabels = target_names, yticklabels = target_names) and is displayed using plt.show().

4. Results

VGG16 is employed as a fixed feature extractor, with the stacked layers subsequently trained on new datasets. Subsequently, to enhance the network performance, certain layers of the base model are unfrozen, allowing them to be updated during training.

In neural network architecture, weights and biases collectively constitute the parameters. Parameters that remain fixed and unaltered during the training process are referred to as non-trainable parameters.

Table 1 delineates the number of trainable and non-trainable parameters for the customized model.

Table 1. Trainable and non-trainable parameters.

Model	Total Parameters	Trainable	Non-Trainable
Customized Model	39,805,690	25,091,002	14,714,688

A customized model is trained and validated by tuning the learning rate parameter and the number of epochs. The testing and validation accuracies of the trained models are shown in Table 2. The total running time is 484.59 s.

Table 2. The testing and validation accuracies of the trained models.

No. of Epoch	Learning Rate	Testing Loss	Testing Accuracy	Validation Loss	Validation Accuracy
Epoch 1/10	15 s 3 s/step	6.2137	0.5000	6.2114	0.5000
Epoch 2/10	16 s 3 s/step	2.0953	0.5217	2.1691	0.5100
Epoch 3/10	15 s 3 s/step	1.7656	0.5669	1.7505	0.5500
Epoch 4/10	16 s 3 s/step	0.8504	0.7587	0.9814	0.7300
Epoch 5/10	15 s 3 s/step	0.9370	0.7054	0.7952	0.7400
Epoch 6/10	16 s 3 s/step	0.6226	0.7319	0.6401	0.7100
Epoch 7/10	15 s 3 s/step	0.6660	0.7668	0.7453	0.7400
Epoch 8/10	15 s 3 s/step	0.5976	0.7720	0.6835	0.7400
Epoch 9/10	20 s 5 s/step	0.5129	0.7398	0.5192	0.7400
Epoch 10/10	16 s 3 s/step	0.6034	0.7305	0.6018	0.7300

From the above table, it is evident that the accuracy rate increases as the number of epochs increases for the same network. Additionally, there exists a relationship between the training accuracy and the learning rate.

The graphs (Figure 8) depicting the training loss, validation loss, training accuracy, and validation accuracy are crucial for evaluating the model's performance during training. A decrease in the training loss signifies improved learning and levels of efficacy for the training data, while a validation loss indicates the model's ability to generalize with new data. The training accuracy shows how well the model classifies or predicts the training data, typically improving over time. The validation accuracy reflects the model's generalization ability for unseen data; stability or an improvement throughout training suggests it has an effective generalization ability and prevents overfitting.

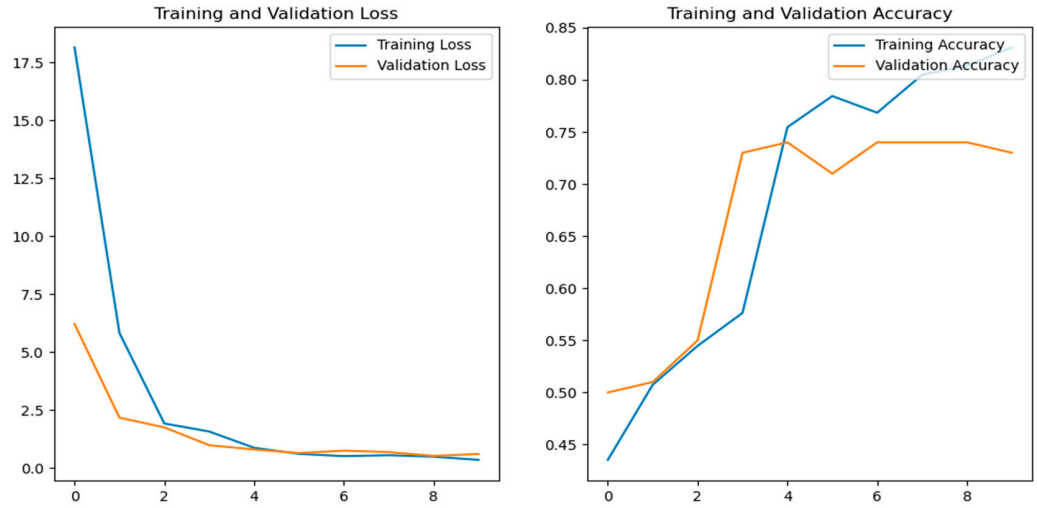


Figure 8. Training loss, validation loss, training accuracy, and validation accuracy versus number of epochs.

The performance of the classification model was evaluated using its precision, recall, F1 score, and accuracy metrics (Table 3). While the overall accuracy of the model was over 70%, slightly surpassing that of making random guesses, there remains room for improvement, suggesting opportunities for enhancing model performance. The evaluation was conducted on an independent test set, ensuring the model's performance was assessed on unseen data, enhancing its reliability in real-world applications. Additionally, macro and weighted averages for its precision, recall, and F1 score were consistent with individual class metrics, providing a comprehensive overview of the model's performance across all categories. The model's performance is relatively balanced between the two classes, but there is room for improvement, particularly in terms of its recall for both classes. These results lay the foundation for the further refinement and optimization of the classification model to enhance its efficacy in diagnosing celiac disease. The findings obtained in this study are a part of the author's doctoral thesis.

Table 3. The performance of the classification model.

	Precision	Recall	F1 Score	Support
Celiac	0.57	0.56	0.52	50
Nonceliac	0.54	0.70	0.61	50
Accuracy			0.55	100
Macro Avg	0.55	0.55	0.54	100
Weighted Avg	0.55	0.55	0.54	100

5. Discussion

This study highlights the potential of transfer learning, particularly the VGG16 model, in the diagnosis of celiac disease through the classification of facial photos. The model achieved a 73% accuracy with sensitivity, recall, and F1 score values of 0.54, 0.56, and 0.52, respectively. While these results are promising, they also highlight challenges and areas for improvement.

The model demonstrated a moderate level of accuracy, demonstrating its ability to identify subtle facial features associated with celiac disease. However, the balanced but moderate levels of sensitivity and recall suggest a need for improvement.

The dataset was limited in size due to variations in sex ratios and dietary compliance levels, which provided variability. This small dataset likely contributed to performance limitations and highlights the need for larger and more diverse datasets. Classifying facial

images, which can be challenging even for experts, posed a significant challenge. Variability in facial features due to irrelevant factors, such as genetics and age, further complicated the model's task.

Transfer learning with VGG16 accelerated the training process and achieved a reasonable performance with a small dataset. Resizing images from 1600 × 2000 to 224 × 224 pixels, as required by models such as VGG16, can introduce artifacts and obscure fine details. While these models can still extract robust features, the loss of fine details in patients' facial photographs can negatively impact model performance. Future work can investigate preserving critical facial details during preprocessing to improve accuracy. Furthermore, since we intend to create a web-based system to which patients can upload images for prediction purposes in our future work, these images may vary in size and may even be low-resolution. Furthermore, we compared a fixed training–test split (using the holdout method) and K-fold cross-validation. The K-fold cross-validation resulted in a lower accuracy due to the difficulty distinguishing between classes, suggesting the need for more data or different modeling approaches. Consequently, we chose the holdout method for its higher accuracy and continued the process.

Future research using a larger dataset is recommended. Increasing the dataset size and ensuring balanced representation can improve the performance of the model by providing a more diverse set of training examples. However, further investigation into specific facial features used by the model may provide insights into the physical symptoms of celiac disease. Facial photo classification can be integrated in the future with other diagnostic tools using a multimodal approach; for example, combining it with serological tests or genetic markers may improve its diagnostic accuracy.

This study demonstrates the feasibility of transfer learning with VGG16 for the diagnosis of celiac disease through facial photo classification, achieving a 73% accuracy. While the model is promising, expanding and diversifying the dataset and improving the model can significantly improve its performance. This research forms the basis for developing a reliable diagnostic tool for celiac disease by utilizing transfer learning in medical image analysis.

6. Conclusions

Looking at existing studies, a thorough literature review reveals a dearth of research focusing on diagnostic aid systems employing transfer learning for classifying photographs of patients with celiac disease, aiming to aid in celiac disease diagnosis or offer guidance for referral to specialists by primary care physicians. The retrained model achieved a 73% accuracy, hinting at its potential as a diagnostic tool for celiac disease identification. However, this also implies that there are opportunities for enhancement, particularly considering the presence of challenging images that pose difficulties even for human classification. In summary, these findings underscore the potential for developing a valuable diagnostic tool for identifying celiac disease using transfer learning techniques in biomedical engineering image classification.

Author Contributions: İ.Z.G.: supervision, study planning, and manuscript editing. E.K.B.: Main manuscript writing, editing, revision, study design, analysis. Y.K.: Data collection. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of Health Sciences University in Turkey (Ethics committee decision number 2022/09-02 and decision date 27 April 2022). Informed consent was obtained from all individual participants included in the study.

Informed Consent Statement: Written informed consent has been obtained from the patient(s) to publish this paper.

Data Availability Statement: The datasets utilized and/or analyzed in this study are not publicly accessible due to their restricted access status.

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

References

1. Caio, G.; Volta, U.; Sapone, A.; Leffler, D.A.; De Giorgio, R.; Catassi, C.; Fasano, A. Celiac disease: A comprehensive current review. *BMC Med.* **2019**, *17*, 142. [CrossRef]
2. Valitutti, F.; Cucchiara, S.; Fasano, A. Celiac Disease and the Microbiome. *Nutrients* **2019**, *11*, 2403. [CrossRef]
3. Elli, L.; Branchi, F.; Tomba, C.; Villalta, D.; Norsa, L.; Ferretti, F.; Roncoroni, L.; Bardella, M.T. Diagnosis of gluten related disorders: Celiac disease, wheat allergy and non-celiac gluten sensitivity. *World J. Gastroenterol.* **2015**, *21*, 7110–7119. [CrossRef]
4. Kayar, Y.; Dertli, R. Association of autoimmune diseases with celiac disease and its risk factors. *Pak. J. Med. Sci.* **2019**, *35*, 1548–1553. [CrossRef]
5. Sadeghi, A.; Rad, N.; Ashtari, S.; Rostami-Nejad, M.; Moradi, A.; Haghbin, M.; Rostami, K.; Volta, U.; Zali, M.R. The value of a biopsy in celiac disease follow up: Assessment of the small bowel after 6 and 24 months treatment with a gluten free diet. *Rev. Esp. Enferm. Dig.* **2019**, *112*, 101–108. [CrossRef] [PubMed]
6. Keskin, E. Clinical Decision Support Systems in Diagnosis of Autoimmune Diseases. In Proceedings of the Bilge Kagan 2nd International Science Congress Barcelona, Barcelona, Spain, 5–7 November 2019; pp. 239–250.
7. Russo, E.; Di Gloria, L.; Cerboneschi, M.; Smeazzetto, S.; Baruzzi, G.P.; Romano, F.; Ramazzotti, M.; Amedei, A. Facial Skin Microbiome: Aging-Related Changes and Exploratory Functional Associations with Host Genetic Factors, a Pilot Study. *Biomedicines* **2023**, *11*, 684. [CrossRef] [PubMed]
8. Sinha, S.; Lin, G.; Ferenczi, K. The skin microbiome and the gut-skin axis. *Clin. Dermatol.* **2021**, *39*, 829–839. [CrossRef] [PubMed]
9. Finizio, M.; Quaremba, G.; Mazzacca, G.; Ciacci, C. Large forehead: A novel sign of undiagnosed coeliac disease. *Dig. Liver Dis.* **2005**, *37*, 659–664. [CrossRef]
10. Caproni, M.; Bonciolini, V.; D'Errico, A.; Antiga, E.; Fabbri, P. Celiac Disease and Dermatologic Manifestations: Many Skin Clue to Unfold Gluten-Sensitive Enteropathy. *Gastroenterol. Res. Pract.* **2012**, *2012*, 952753. [CrossRef]
11. Krzywicka, B.; Herman, K.; Kowalczyk-Zając, M.; Pytrus, T. Celiac disease and its impact on the oral health status—Review of the literature. *Adv. Clin. Exp. Med.* **2014**, *23*, 675–681. [CrossRef]
12. Sernicola, A.; Alaibac, M. Editorial: Cutaneous manifestations of systemic disease. *Front. Med.* **2023**, *10*, 1236570. [CrossRef] [PubMed]
13. Durazzo, M.; Ferro, A.; Brascugli, I.; Mattivi, S.; Fagoonee, S.; Pellicano, R. Extra-Intestinal Manifestations of Celiac Disease: What Should We Know in 2022? *J. Clin. Med.* **2022**, *11*, 258. [CrossRef] [PubMed]
14. Sali, R.; Adewole, S.; Ehsan, L.; Denson, L.A.; Kelly, P.; Amadi, B.C.; Holtz, L.; Ali, S.A.; Moore, S.R.; Syed, S.; et al. Hierarchical Deep Convolutional Neural Networks for Multi-category Diagnosis of Gastrointestinal Disorders on Histopathological Images. In Proceedings of the 2020 IEEE International Conference on Healthcare Informatics (ICHI), Oldenburg, Germany, 30 November–3 December 2020.
15. Singh, A.; Kisku, D.R. Detection of Rare Genetic Diseases using Facial 2D Images with Transfer Learning. In Proceedings of the 2018 8th International Symposium on Embedded Computing and System Design, ISED 2018, Cochin, India, 13–15 December 2018; pp. 26–30. [CrossRef]
16. Ismail, A.R.; Nisa, S.Q.; Shaharuddin, S.A.; Masni, S.I.; Amin, S.A.S. Utilising VGG-16 of Convolutional Neural Network for Medical Image Classification. *Int. J. Perceptive Cogn. Comput.* **2024**, *10*, 113–118. [CrossRef]
17. Cireşan, D.C.; Meier, U.; Masci, J.; Gambardella, L.M.; Schmidhuber, J. Flexible, High Performance Convolutional Neural Networks for Image Classification. In Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Spain, 16–22 July 2011; pp. 1237–1242.
18. Qiu, J.; Wu, Q.; Ding, G.; Xu, Y.; Feng, S. A survey of machine learning for big data processing. *EURASIP J. Adv. Signal Process.* **2016**, *2016*, 67. [CrossRef]
19. Zheng, J.; Shen, F.; Fan, H.; Zhao, J. An online incremental learning support vector machine for large-scale data. *Neural Comput. Appl.* **2012**, *22*, 1023–1035. [CrossRef]
20. Lundervold, A.S.; Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. *Z. Fur Med. Phys.* **2018**, *29*, 102–127. [CrossRef] [PubMed]
21. Shrestha, A.; Mahmood, A. Review of deep learning algorithms and architectures. *IEEE Access* **2019**, *7*, 53040–53065. [CrossRef]
22. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
23. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018. [CrossRef]
24. Bengio, Y. Deep Learning of Representations for Unsupervised and Transfer Learning. 2012. Available online: http://www.causality.inf.ethz.ch/unsupervised-learning.php (accessed on 8 July 2024).
25. Gadermayr, M.; Wimmer, G.; Kogler, H.; Vécsei, A.; Merhof, D.; Uhl, A. Automated classification of celiac disease during upper endoscopy: Status quo and quo vadis. *Comput. Biol. Med.* **2018**, *102*, 221–226. [CrossRef]

26. Seguí, S.; Drozdzal, M.; Pascual, G.; Radeva, P.; Malagelada, C.; Azpiroz, F.; Vitrià, J. Generic feature learning for wireless capsule endoscopy analysis. *Comput. Biol. Med.* **2016**, *79*, 163–172. [CrossRef]
27. Wang, X.; Qian, H.; Ciaccio, E.J.; Lewis, S.K.; Bhagat, G.; Green, P.H.; Xu, S.; Huang, L.; Gao, R.; Liu, Y. Celiac disease diagnosis from videocapsule endoscopy images with residual learning and deep feature extraction. *Comput. Methods Programs Biomed.* **2019**, *187*, 105236. [CrossRef] [PubMed]
28. Amirkhani, A.; Mosavi, M.R.; Mohammadi, K.; Papageorgiou, E.I. A novel hybrid method based on fuzzy cognitive maps and fuzzy clustering algorithms for grading celiac disease. *Neural Comput. Appl.* **2018**, *30*, 1573–1588. [CrossRef]
29. Vimal, C.; Shirivastava, N. Face and Face-mask Detection System using VGG-16 Architecture based on Convolutional Neural Network. *Int. J. Comput. Appl.* **2022**, *183*, 16–21. [CrossRef]
30. Hands-On Transfer Learning with Keras and the VGG16 Model—LearnDataSci. Available online: https://www.learndatasci.com/tutorials/hands-on-transfer-learning-keras/ (accessed on 3 April 2024).
31. Akhand, M.A.H.; Roy, S.; Siddique, N.; Kamal, A.S.; Shimamura, T. Facial emotion recognition using transfer learning in the deep CNN. *Electronics* **2021**, *10*, 1036. [CrossRef]
32. Su, Z.; Liang, B.; Shi, F.; Gelfond, J.; Šegalo, S.; Wang, J.; Jia, P.; Hao, X. Deep learning-based facial image analysis in medical research: A systematic review protocol. *BMJ Open* **2021**, *11*, e047549. [CrossRef]
33. Sharma, R.; Scholar, M.T. Dataset to Model: Optimization of Image Classification with Neural Networks. 2023. Available online: http://www.ijeast.com (accessed on 8 July 2024).

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.