*Article*

# Deep-Learning-Based System for Assisting People with Alzheimer's Disease

Dan Munteanu [1], Catalina Bejan [1], Nicoleta Munteanu [1], Cristina Zamfir [2], Mile Vasić [3], Stefan-Mihai Petrea [2,4,*] and Dragos Cristea [2,*]

[1] Faculty of Automation, Computer Sciences, Electronics and Electrical Engineering, "Dunarea de Jos" University of Galati, 111 Domnească Street, 800201 Galaţi, Romania

[2] Faculty of Economics and Business Administration, "Dunarea de Jos" University of Galati, 47 Domnească Street, 800201 Galaţi, Romania

[3] European Marketing and Management Association, Knežopoljska 5, 78000 Banja Luka, Bosnia and Herzegovina

[4] Food Science, Food Engineering, Biotechnology and Aquaculture (SAIABA), Faculty of Food Science and Engineering, "Dunarea de Jos" University of Galati, 47 Domnească Street, 800201 Galaţi, Romania

* Correspondence: stefan.petrea@ugal.ro (S.-M.P.); dragoscristea@yahoo.com (D.C.)

**Abstract:** People with Alzheimer's disease are at risk of malnutrition, overeating, and dehydration because short-term memory loss can lead to confusion. They need a caregiver to ensure they adhere to the main meals of the day and are properly hydrated. The purpose of this paper is to present an artificial intelligence system prototype based on deep learning algorithms aiming to help Alzheimer's disease patients regain part of the normal individual comfort and independence. The proposed system uses artificial intelligence to recognize human activity in video, being able to identify the times when the monitored person is feeding or hydrating, reminding them using audio messages that they forgot to eat or drink or that they ate too much. It also allows for the remote supervision and management of the nutrition program by a caregiver. The paper includes the study, search, training, and use of models and algorithms specific to the field of deep learning applied to computer vision to classify images, detect objects in images, and recognize human activity video streams. This research shows that, even using standard computational hardware, neural networks' training provided good predictive capabilities for the models (image classification 96%, object detection 74%, and activity analysis 78%), with the training performed in less than 48 h, while the resulting model deployed on the portable development board offered fast response times—that is, two seconds. Thus, the current study emphasizes the importance of artificial intelligence used in helping both people with Alzheimer's disease and their caregivers, filling an empty slot in the smart assistance software domain.

**Keywords:** Alzheimer's disease; artificial intelligence; computer vision; surveillance system; image classification; deep learning; transfer learning; object detection; human activity recognition in video

## 1. Introduction

Just as the human body ages over time, so does the brain. Old age is the highest risk factor for Alzheimer's disease, with most cases developing after the age of 65. However, there have been incipient cases of the disease even below this age, proving that it not only affects the elderly, but also is not a natural stage of aging [1].

Dementia is a general term used to define a set of abnormalities that occur in the onset of the disease such as impaired memory and thinking skills. Alzheimer's disease occurs in 60–80% of cases of dementia, causing a gradual loss of the ability to function independently [1].

There are 50 million people suffering from dementia worldwide and it is estimated that this number will double every 2 decades, reaching 152 million by 2050 [2]. This alarming

increase in the number of patients has led to scientists racing against the clock to find effective treatments to prevent and alleviate the symptoms of the disease [3].

Being a progressive disease, in the initial phase, it manifests in slight memory loss, but worsens over time to the level where the affected person can no longer communicate in words or respond to environmental stimuli [4].

Alzheimer's disease can cause an eating disorder—patients either forget to eat, causing malnutrition, or eat too many times a day, leading to overeating [5].

For people with Alzheimer's disease, the risk of dehydration or unhealthy eating is much higher; they even forget to eat or drink water, regardless of having a feeling of thirst or hunger. Meanwhile, patients are prone to eat more than necessary or more often as a result of the confusion they feel because they cannot rely on short-term memory [6].

The possibility of living independently thus becomes almost impossible and the need for a caregiver is essential.

Thus, there is a need for a system that monitors a patient's activities to track the times when he/she eats and drinks, and warns both him and his caregiver of the crucial moment at which the patient should drink water, eat, or stop eating in order to avoid dehydration, malnutrition, or overnutrition.

The existing systems, designed to help people with dementia and Alzheimer's disease, consist of person tracking systems, reminders of activities such as taking medication, and video surveillance systems to track the patient's activities by the caregiver. As examples, the following can be mentioned:

- Reminder applications: (a) "Pill Reminder" (iOS) is a reminder application that alerts the patient that he must take his medication at a fixed time. These alarms are set by a caregiver and the app allows the patient to log in manually when the pill was administered. (b) "It's Done!" (iOS) is a reminder application in which the caregiver can set a series of activities to be performed in a day and the patient can check them as satisfied after performing them [7].
- Surveillance systems: the existing surveillance systems allow real-time video streaming to the caregiver, but to be effective, the caregiver needs to constantly monitor the patient's activity through them, that is, to watch the video streaming continuously. "Foscam" surveillance cameras allow this, as well as the possibility of audio communication between the patient and the caregiver. "Y-cam" cameras allow motion-triggered surveillance. Other real-time surveillance cameras include Netgear Arlo and Baby Cam [8].

Currently, there are few artificial intelligence-based solutions available on the market targeting an active real-time assistance of AD patients for critical aspects like nutrition or hydration. Most of the applications refer to memory stimulation, to-do lists, and perimeter surveillance. The following solutions were identified: (a) Constant Therapy, an app for smartphones and tablets that offers cognitive, language, and speech therapy, designed to support patients with Alzheimer's disease and dementia, as well as those recovering from brain injuries including stroke. (b) Alz Calls, a chatbot designed for patients who repeatedly ask for their family, struggle with transitions to new environments, or need social interaction. Family members can record their voice, add a photo that will pop up for the patient to recognize, and answer frequent questions so that the patient can have an interactive conversation when the caregiver is not available to talk. (c) AmuseIT is an app designed to promote conversation. It contains over 1000 simple quiz questions with a strong visual component. (d) Timeless, an app for Alzheimer's patients to remember events, stay connected and engaged with friends and family, as well as to recognize people through artificial-intelligence-based facial recognition technology. (e) MindMate, an all-in-one resource for dementia patients. This app was developed based on the latest research on dementia and helps curtail some of the effects of cognitive decline using brain-engaging games, nutrition advice, and exercise programs. (f) MemoClock, an app that connects a caregiver with their friend or relative who has dementia and allows them to send reminders directly to the other person's phone. For daily tasks, repeat reminders can be set. (g) It's

Done!, an application with 40 daily tasks to help dementia patients remember to-dos. App users can check off their tasks as they do them and can also create their own task list to remember unique tasks. (h) Ode, an app that releases three natural aromas of food to coincide with mealtimes for stimulating patients' appetite. (i) Human Link, a package consisting of a watch that has an integrated SIM, a watch that is worn by the person with Alzheimer's disease, and an application that is installed on the family's phones. The watch also has a phone function, through which stored preferential numbers can be contacted. It is also possible to establish a perimeter in which the person carries out his activity, areas called safety zones. If the patient comes out of this perimeter, the family will be automatically notified.

From the systems and applications designed to help people with Alzheimer's, there is currently no application or system that monitors the patient's activities to prevent dehydration, malnutrition, or overeating. At the same time, there is no application that allows caregivers to give up the constant monitoring of the activities of the patients they care for because there is no system with automatic recognition by artificial intelligence of the activities designed for them.

The developed surveillance systems, developed for both people with Alzheimer disease and their caretakers, allows the monitoring and recognition of human activity; the real-time transmission of the recording; the saving of the moments when the supervised person eats or drinks in a database; and the notification of the patient and the caregiver when the patient should eat, drink, or stop eating.

Thus, the person with Alzheimer's disease can regain his independence by not having to live in the same house with his caregivers. At the same time, the caregiver enjoys security in terms of the needs and health of the patient by the fact that it is no longer necessary to constantly monitor his activity; this will be handled by the surveillance system based on artificial vision. Moreover, if the caregiver wants to monitor the patient's activities, he can do it remotely by simply connecting to the internet from any device that has access to an internet browser.

This paper includes the study, search, training, and use of models and algorithms specific to the field of artificial vision to classify images, detect objects in images, and recognize human activity in video.

The current research contributes to the Healthcare 4.0 paradigm, presented in [9], a paradigm involving the realization of data-driven and patient-centric health systems wherein advanced sensors can be deployed to provide personalized assistance. As presented by the authors, mentally affected patients from diseases like Alzheimer can be assisted using sophisticated algorithms and enabling technologies. Their research targets sensory movement data as well as emotion detection. Another study [10] attempts to develop a platform to support patients who suffer from face perception impairment with an assistive intelligence device. As the authors mentioned, up to now, the development of wearable IoT devices has been hindered by the lack of their computational power, but now, the portable computational capabilities of the proposed system were enhanced so much that real-time operations are possible. In this context, the current study provides several major contributions to the overall Healthcare 4.0 environment: (a) it targets a specific and vital activity (food habits) of Alzheimer patients, an activity that was not covered extensively by other studies; (b) the proposed algorithms could be deployed in specific systems not necessarily for Alzheimer disease; and (c) it proves that, indeed, the current portable hardware is able to run almost real-time artificial intelligence inferences.

## 2. Materials and Methods

Artificial intelligence has reached its current level of development thanks to the constant desire of experts in the field to build machines that can accomplish everything a human being can do.

This field is constantly evolving, already being helpful in people's daily lives (through personal assistants on smartphones such as Siri, Alexa, Google Assistant) and even exceed-

ing human physical capabilities in some situations (exploitation of harmful chemicals and work in large production plants) [11].

Artificial intelligence and its applicability in everyday activities have especially progressed in the last decade. The purpose of the development of artificial intelligence is to create machines that can perceive, understand, interpret, determine, rationalize, recognize, and react to environmental stimuli, just like a human being [11].

### 2.1. Modern Artificial Intelligence

The 2012 discoveries (including a neural network built by researcher Geoffrey Hinton and his team, with which they won an international artificial vision competition) [12] led to today's advances in artificial vision, language recognition, and self-driving cars. These factors have changed the way artificial intelligence is viewed today [11].

The subfields of artificial intelligence include the following:

- Natural language processing (deals with the analysis and understanding of natural language information in documents);
- Robotics (deals with the design, programming, and manufacture of robots);
- Machine learning (the ability of a machine to learn) and deep learning (sub-domain of machine learning);
- Expert systems (solve complex problems by analyzing a knowledge base containing rules and generating a conclusion);
- Speech recognition (deals with speech and voice recognition);
- Process automation (deals with automating robots in combination with artificial intelligence);
- Artificial vision (the ability of machines to see and understand as well as interpret what they see).

It is often necessary to use several subdomains to build a project based on artificial intelligence, as they are interconnected.

### 2.2. Artificial Neural Networks

The method using neural networks was developed according to the biological model of the human brain. This approach involves mimicking the networks of neurons found in the brains of humans and animals, being the only form of intelligence we know. Artificial neural networks can learn from experience. This assumes that, if an artificial neural network has previously received several examples of images containing food plates, then it can identify patterns from which to deduce whether an image never seen before resembles those labeled as food plates [12,13].

Artificial neural networks require high computational power, but also large datasets from which to learn [12]. Datasets must be as large as possible to provide a real perspective of the object with which it can be identified in the three-dimensional world in which we live from any angle and at any distance as captured in the digital image (or video file).

During the previous few years, artificial intelligence (AI) revolutionized the neurology practice. Currently, AI can interpret electroencephalograms (EEGs) to prognosticate coma, classify neurodegenerative diseases based on gait and handwriting, diagnose a stroke from CT/MRI scans, detect seizure well before ictus, detect papilledema and diabetic retinopathy from retinal scans, and predict the conversion of mild cognitive impairment to Alzheimer's dementia. Even clinical practice could possibly change in the near future to accommodate AI as a complementary tool [14].

As emphasized in [15], the recent growth in open data-sharing initiatives collecting lifestyle, clinical, and biological data from Alzheimer's patients has provided a potentially unlimited amount of information about the disease, far exceeding the human ability to make sense of it. In this context, AI offers a variety of methods to analyze large and complex data for improving knowledge in the Alzheimer's disease field and to provide aid in the development of computer-aided diagnosis tools for AD diagnosis and personalized therapies.

## 2.3. Datasets

With the advancement of technology, computing power has become sufficient for the use of artificial neural networks, and datasets have become larger and larger, with many examples needed to compile the experience from which the program learns. In the case of image or video datasets, it has been possible to increase the amount of data because of the fact that most people have a device with which they can take pictures and record videos (e.g., smartphone, camera, and webcam). The collection of the respective digital data is possible because of the accessibility to the internet and to the social networks, within which different photos or videos can be uploaded.

## 2.4. Deep Learning

Even though computational power and the size of datasets are no longer an impediment, artificial neural networks required many layers of neurons to be usable in a real-life application, but they could not be effectively trained as they were added to the network. This has changed with the advent of deep learning through the discoveries made by researcher Geoffrey Hinton. It is now possible to effectively train the layers of neurons, making considerable progress for artificial vision, as well as for voice recognition [12].

## 2.5. Convolutional Neural Networks

Convolutional neural networks are developed for processing images or similar 2D/3D images and videos. What sets them apart from simple neural networks is that they are made up of convolutional layers and use pooling layers for image compression. To return the result, these networks use fully connected layers (all neurons are connected).

Applications of interest in the field of artificial vision:

- Image classification—involves analyzing the entire image and classifying it into a single category;
- Detection of objects in images—involves identifying and locating one or more objects, framing them in an image;
- Image segmentation—involves identifying and locating the pixels that make up objects in an image;
- Video activity recognition—involves analyzing an entire video and classifying it into a single category;
- Video activity location—involves identifying and locating one or more actions in a video.

According to a recent study [16], deep learning can be successfully used to diagnose diseases such as skin lesions, brain disease, and chronic disease. Thus, tear film breakup time-based dry eye disease detection using a convolutional neural network was performed in a previous study [17] based on 30 videos collected from a private eye hospital. Fourteen training videos (70% of the total number of videos available) are converted into 600 frames to train the convolutional neural network, and six videos are converted into 100 frames for testing while defining four classes: normal, breakup, blink, or noise. Thus, the authors [17] managed to obtain an accuracy of 83%, 87% sensitivity, 89% specificity, and 90% correlation coefficient.

Furthermore, other authors [18] proved that the combination of deep learning and optical coherence tomography (OCT) technologies has allowed reliable detection of retinal diseases and improved diagnosis. Some researchers [19] consider that advanced deep learning techniques could successfully replace the established image analysis methods, as this trend has been observed within other medical fields such as gastrointestinal diseases and radiology [20,21].

Another study [22] proposes a solution for evaluating the meibomian gland atrophy by extracting features from meibography images with a special type of unsupervised CNN before the application of a K-nearest neighbors model to allocate a meiboscore.

Some authors [23] propose a novel hybrid intelligent diagnosis framework deep fused features-based reliable optimized networks, which fuses the saliency maps and convo-

lutional layers for better segmentation (reduction of overfitting problem) and feature extraction that are used to train the ant-lion optimized single feedforward networks; additionally, hyperparameters of pretrained networks are tuned by the whale optimization algorithm to improve the prediction accuracy by sacrificing the complexity. The obtained results reveal a 98.5% accuracy, 99.0% precision, 98.8% recall, and 99.1% F1-score, outperforming the DenseNet, AlexNet, Resnets-50, Resnets-100, VGG-16, and Inception models. However, other authors [24] designed a controlled CNN classifier for patients with lung cancer to detect potential adenocarcinoma and squamous cell carcinoma and recorded a low predictive performance of 71%.

### 2.6. Image Classification and Transfer Learning

To classify the images, the development of convolutional neural networks was carried out from scratch, as well as the use of the technique of learning through knowledge transfer (transfer learning).

A convolutional neural network was modeled and built using the VGG-16 architecture for image classification and previously trained on the ImageNet dataset.

The choice of this architecture and the binary classification for food identification (food class) was made after the training of several models using both a network architecture built from scratch and knowledge transfer on models such as MobileNet, InceptionV3, and VGG16, as well as different data sets and different classes for identifying images of interest (those containing food).

### 2.7. The Dataset

The datasets used to drive the image classification models are Food-101 [25], Furniture Detector [26], and the IMDB-WIKI dataset [27–29].

As it is desirable to classify images as food or non-food, the datasets mentioned above include either images of food or food images that are frequently encountered (furniture and people).

## 3. Related Works

Alzheimer's disease has been subject to various studies that employ supervised learning techniques such as deep neural network (DNN), convolutional neural network (CNN), and recurrent neural network (RNN) [30,31].

Thus, some authors [32] developed a nonlinear metric learning method to improve biomarker identification for Alzheimer's disease and used a deep-network-based feature fusion strategy through a stacked denoising sparse auto-encoder to integrate cross-sectional and longitudinal features estimated from magnetic resonance brain images. Other authors [33] propose a classification method of Alzheimer's disease based on structural modalities using the deep belief network (DBN), which operates in terms of predictive models. Some studies [34] proposed an Alzheimer's disease stage detection system, based on deep features, thus using a pre-trained AlexNet model, transferring the initial layers from the pre-trained model, and extracting the deep features from the convolutional neural network (CNN).

The identification and classification of Alzheimer's disease when limited annotations are available was performed by some authors [35] using a weakly supervised learning-based deep learning (DL) framework (ADGNET) consisting of a backbone network with an attention mechanism and a task network for simultaneous image classification and image reconstruction.

A recent study [36] proposes the use of DenseCNN2, a deep convolutional network model for Alzheimer's disease classification, by incorporating global shape representations along with hippocampus segmentation.

After performing a holistic approach related to Alzheimer's disease diagnosis, some authors [37] revealed that, in terms of classification methods, CNN offers the best performance, compared with other deep models, while others [38] revealed that deep learning

techniques outperform conventional machine learning techniques in Alzheimer's disease diagnosis, while deep neural networks outperform the shallow neural network architecture.

Recurrent neural network was used recently [39] in an automatic assessment system for Alzheimer's disease based on speech. Moreover, other authors employed a transfer-learning approach to improve automatic Alzheimer's disease prediction using a relatively small, targeted speech dataset without using expert-defined linguistic features [40].

It can be concluded that most of the supervised learning techniques were used to recognize/detect/predict the early symptoms of Alzheimer's disease in an efficient manner.

However, we should also aim to upregulate personalized assistance for Alzheimer's disease patients by using modern AI techniques based on computer vision. A recent study [9] proposed a deep-learning-based Internet of Health framework for the assistance of Alzheimer's disease patients that includes both an ensemble approach for abnormality tracking and an IoT-based assistance mechanism that can inform the patients about an incorrect decision and uses reinforcing and supportive assistance to support the correct decision.

As previously mentioned, activities such as eating and drinking represent a challenge for Alzheimer's disease patients; therefore, a deep-learning-based Internet of Health framework focused on supporting both the patient and caregiver, which could detect food and drinks and, therefore, monitor human activities related to eating and drinking to correct the behavior, could be an efficient solution for real-time personalized assistance.

According to some authors [41], the importance and impact of food in people's life required the development of new innovative computer vision methods for automatic food detection and recognition for automatic harvesting; food quality assessment for industry and food classification; and retrieval for food intake monitoring, namely for patients with diseases related to this aspect, such as obesity, diabetes, or Alzheimer's disease.

Thus, some authors [42] used segmentation on food images and then extracted multiple different features from the obtained segments followed by the SOM (Self-Organizing Feature Map) application to teach the model and assess the quality of meals and, later, applied artificial neural network (ANN) to improve classification performance.

Some authors [43] developed a monitoring system for assessing food intake using real-time inputs from a smartphone camera, while calorie count by applying a deep learning approach using a food images database was tested successfully in another study by combining two convolutional neural networks (CNNs) for meal detection and food recognition [43].

A recent study [44] aimed to use an AI-based monitoring system for nutrient intake by hospitalised patients, based on RGB-D image pairs captured before and after the meal, using multi-task fully convolutional network for image segmentation.

Therefore, it can be concluded that food detection and classification using deep learning methods can be successfully performed with high accuracy. This can in turn represent the basis for developing new patient health assistance tools for supporting correct decisions, useful especially for Alzheimer's disease patients.

Nevertheless, according to other authors [45], one limitation of applying deep learning within the Alzheimer's disease area is that it is difficult to modify potential bias in the network when the complexity is too great to guarantee transparency and reproducibility. However, this could be solved through the accumulation of large-scale imaging data and by studying the relationships between deep learning and features. Moreover, when used for Alzheimer's disease detection, according to other studies [46], deep learning can be limited as finding the best combination of different biomarkers requires a detailed explanation of the depth model and benchmarking platform.

Moreover, when it comes to deep learning food and drink intake recognition, most studies rely on experimental work undertaken in extremely constrained environments and scenarios, a fact that should be considered as the main limitation.

Moreover, according to other studies [17], the moderately sized datasets used for training because of the lack of public datasets could represent a serious limitation in the

case of deep learning applications for disease identification and the development of patient support tools.

## 4. Results

### 4.1. Image Classification—Food Versus Non-Food

Developing a deep-learning-based system for assisting people with Alzheimer's disease involved the usage of three types of AI models: image classification, object detection, and human activity recognition.

The image classification analysis was needed for providing an important system functionality, that is, the ability to discern between food and non-food products. The assumption of the study was that, by using the current deep learning algorithms, it is possible to obtain models with very good classification capabilities (above 80%). As further presented, the developed model reached 96% percent accuracy, high enough to be considered of good use.

The Jupyter Notebook environment and Python programming language were used for image classification modeling. The software library dependencies used were NumPy, used for numerical operations (including mathematical functions for working with lists and multidimensional matrices); Keras, for modeling the artificial neural network; Glob, for obtaining file paths from the directory specified (getting the list of paths of all images in the dataset); Matplotlib, for viewing graphs; and Scikit-learn (also known as sklearn), for calculating the accuracy of the model on the test subset and creating the confusion matrix.

Each class within the training subset contains 1000 images and each class in the test subset contains 250 images, i.e., the dataset is divided into 80% data for training and 20% data for testing—see Figures 1 and 2.

```
▼ 📁 food_notfood_2classes_dataset
    ▼ 📁 test
        ▶ 📁 food
        ▶ 📁 notfood
    ▼ 📁 train
        ▶ 📁 food
        ▶ 📁 notfood
```
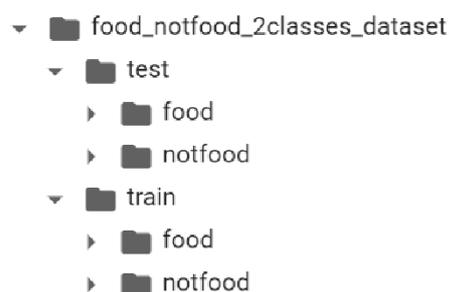
**Figure 1.** Structure of the binary classification dataset.

```
Number of images per class:
class           train   test
food:           1000    250

notfood:        1000    250

Total:          2000    500
```

**Figure 2.** Binary classification dataset divided into 80% training data and 20% testing data.

Considering this method, for a faster model training based on a small dataset to obtain good accuracy on the test set, the knowledge transfer technique was used to train the model. Thus, the base model used to train the new model was the VGG16 convolutional neural network model—see Figure 3. This model was already trained over a period of several weeks on the ImageNet dataset, achieving 92.7% accuracy on the ImageNet dataset (having 14 million images), ranking in the top 5 models tested on the ImageNet dataset [47].
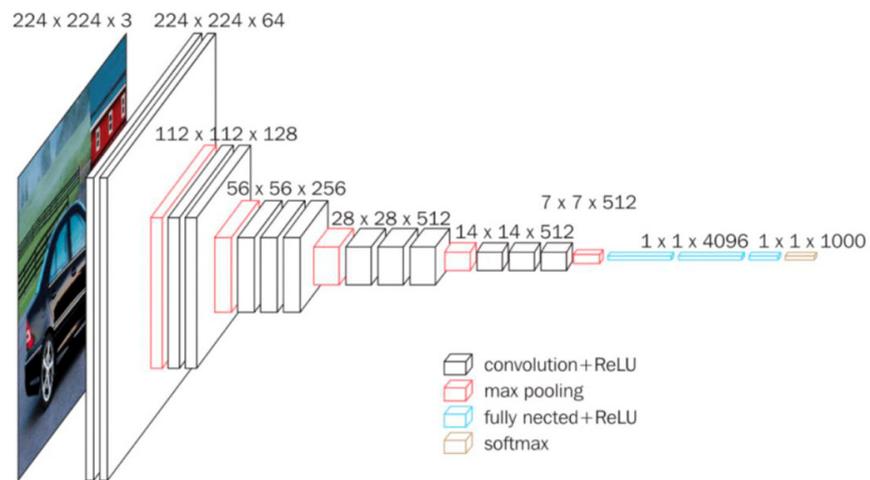
**Figure 3.** Convolutional blocks in VGG16 architecture [48].

As described in Figure 3, the VGG16 architecture consists of five convolutional blocks, with the first two blocks being composed of two convolutional layers and a max pooling layer. The next three convolutional blocks consist of three convolutional layers and a max pooling layer.

In Figure 4, the summary graphs of the training process over 30 epochs are visible; one can notice how the accuracy of the model on the training and test set increased during training and the loss decreased during training, resulting in a successful training of the model, with an accuracy score of 96.8% after training the model.
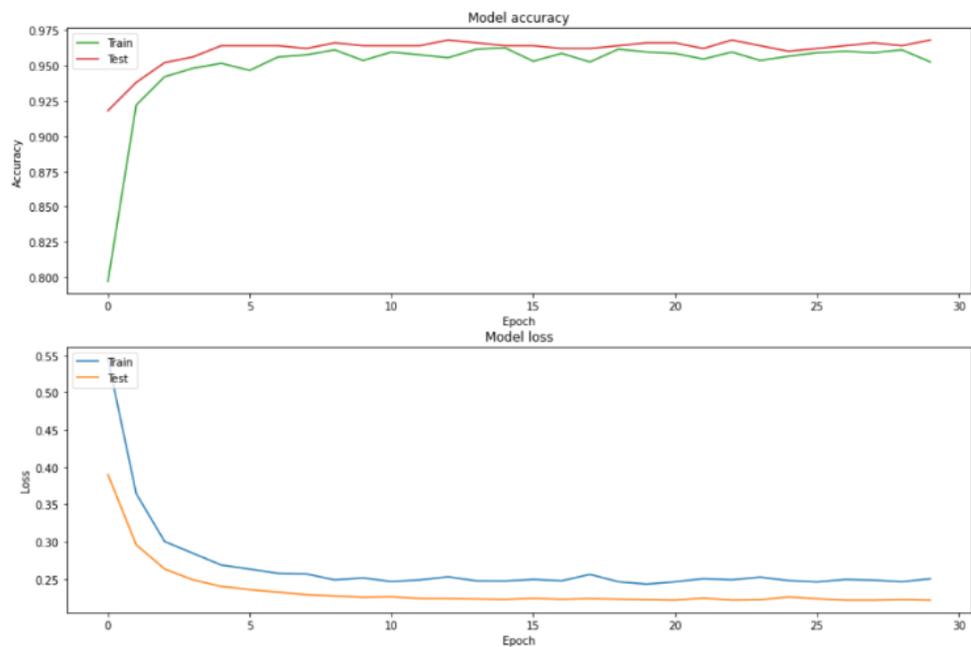


**Figure 4.** Graphs of the training process.

To verify the constructed convolutional neural network, ten random images from the test set and their predictions were inferred.

Figure 5 displays ten randomly selected images along with their correct label, model prediction, and score. Of the ten images, all were correctly classified, with the predictions being in line with their actual label.
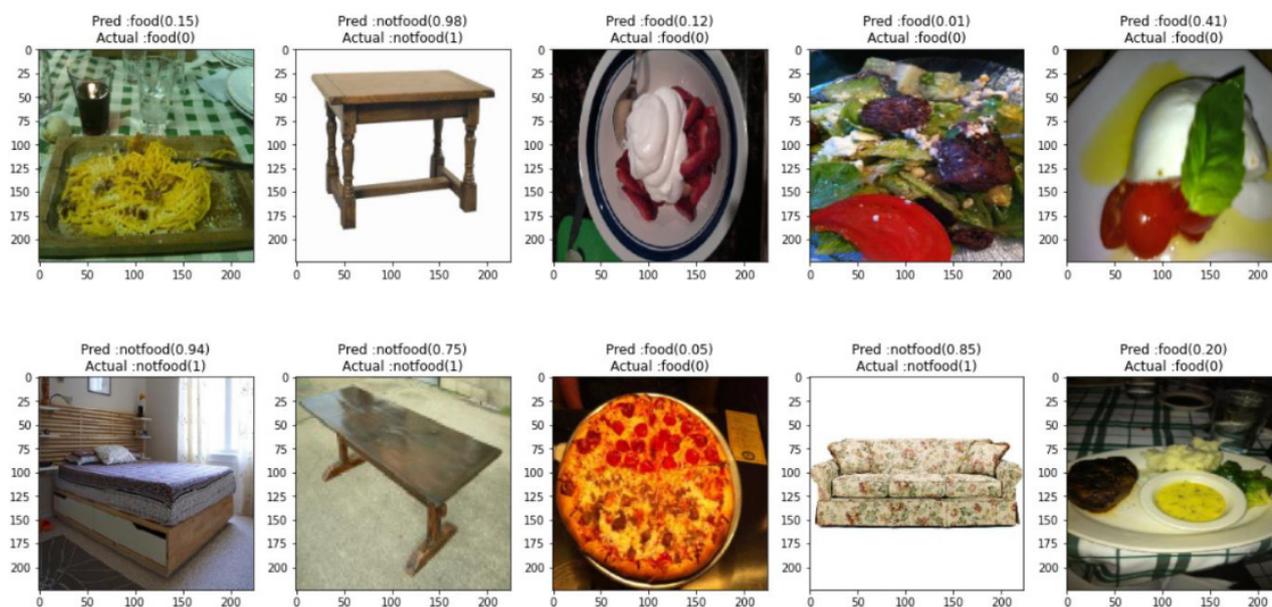
**Figure 5.** View of the correct images, predictions, and label.

We further investigated the model accuracy through a confusion matrix (Figure 6). It was observed that, based on 500 tested images, the following results were obtained: 241 true-positive (the model correctly predicted the positive class—the searched one = "food"), 243 true-negative (the model correctly predicted the negative class—not food), 7 false-positive (model incorrectly predicted the positive class), and 9 false-negative (model incorrectly predicted the negative class) [49].
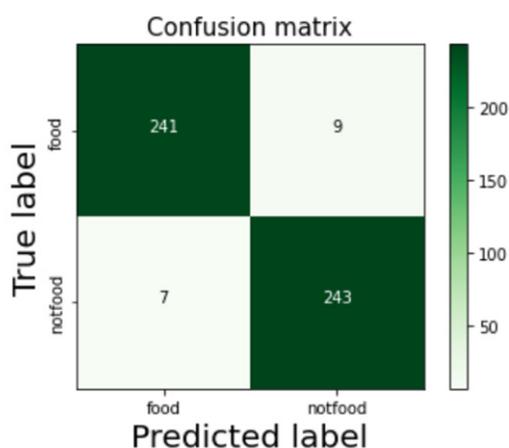


**Figure 6.** Confusion matrix.

From the confusion matrix, it is observable that the number of false predictions is low compared with the total number of predictions made, indicating a high-performance trained model.

### 4.2. Detection of Objects in Images

The object detection algorithms provide the system with the capability of not only classifying food and non-food objects, but also precisely recognizing the name of the food as well as several other object names, like forks or knives. The assumption is that such a system could offer fast (real-time) and good prediction capabilities as long as it is trained on a sufficient amount of data. This research presents a prototype model for this type of functionality; still, for an end product, the dataset needed for object detection should be

enriched with lots of images corresponding to every type of food or object that needs to be identified. As there is such a diversity in food products, hence the need for a highly extended dataset. Still, with proper hardware equipment (high-end video processors), the training of the models could be performed quickly.

Detecting objects in images involves recognizing and locating objects within the image. The objects of interest identified will be framed in a frame drawn on the image on which the object detection process was performed (Figure 7).

For the detection of objects in the images, a convolutional neural network was trained by the knowledge transfer technique based on the SSD ResNet50 V1 FPN 640 × 640 (RetinaNet50) architecture. At the same time, a pre-trained faster R-CNN ResNet50 V1 640 × 640 model (with a high-performance score on the Coco2017 dataset) was also tested.

The two pre-trained models can be found, free to use, in the collection of models called "Tensorflow 2 Detection Model Zoo" [50].

The annotation of an own data set, consisting of images containing objects of interest, was also carried out. The images were searched and selected manually and then annotated using the LabelImg tool (Figure 7) [51].
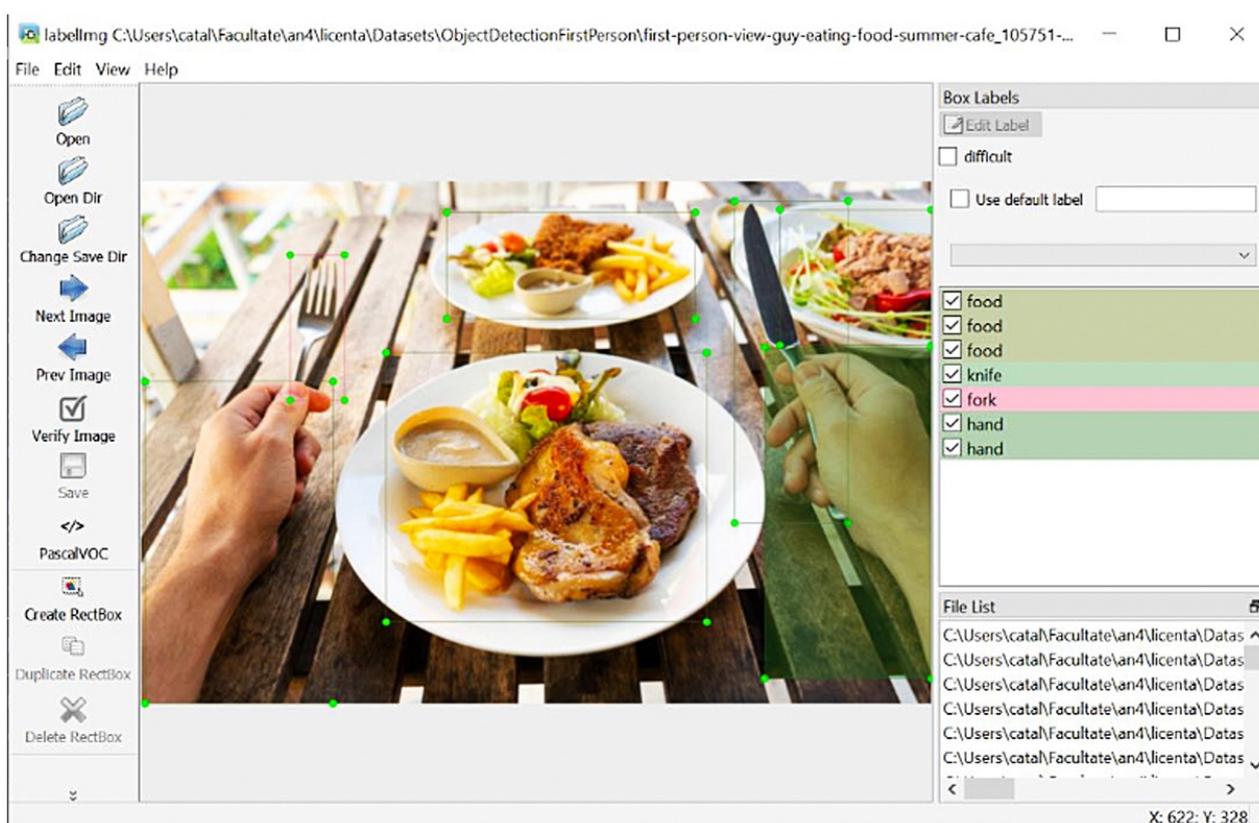


**Figure 7.** Annotating images [52].

Object detection in images involves recognizing and locating objects within the image. The identified objects of interest will be framed in a frame drawn on the image on which the object detection process was performed (see Figure 8). The label of the object found will be mentioned above each border, in addition to the probability score of the prediction made. The model generates the coordinates of the points within which the object of interest is identified and, based on these coordinates, the corresponding border is drawn over the original image.
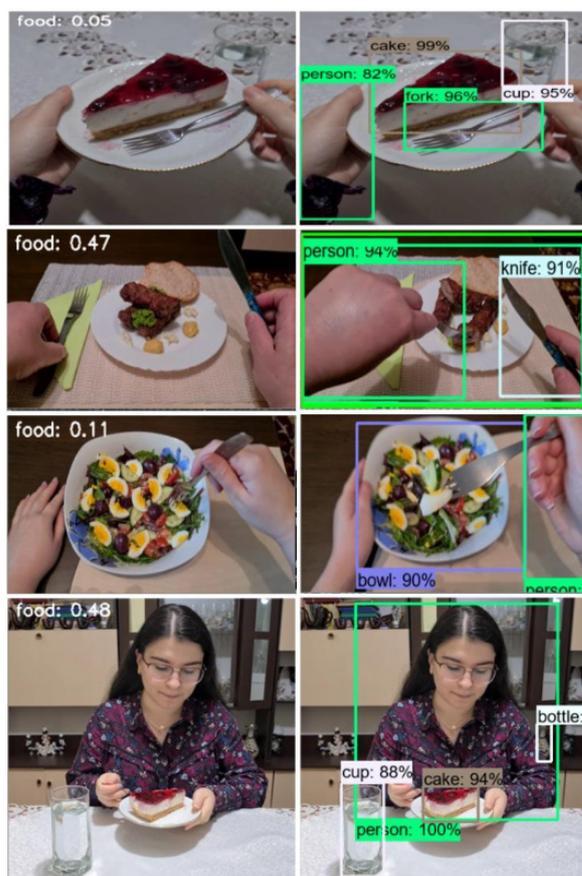
**Figure 8.** The difference between image classification and object detection.

Following the annotation process, a map of object labels in the dataset was created (see Figure 9), having as entries in this file all the classes from the dataset on which the training was performed.

```
 1   item {
 2         id: 1
 3         name: 'apple'
 4   }
 5   item {
 6         id: 2
 7         name: 'orange'
 8   }
 9   item {
10         id: 3
11         name: 'banana'
12   }
```

**Figure 9.** Map of object class labels.

The ResNet50 V1 FPN 640 × 640 (RetinaNet50) SSD model, used for the detection of objects in the images, displays the ResNet50 network architecture containing 50 convolutional layers (see Figure 10). The size of the supported images on the input layer is 640 × 640 pixels, containing 50 convolutional layers, 1 max pooling layer, and 1 average pooling layer (for image compression). The last layer is a fully connected layer (where all neurons are connected), which is represented by the SoftMax function specific to the classification of the 1000 class categories from the ImageNet dataset on which the model with this architecture was trained [53].
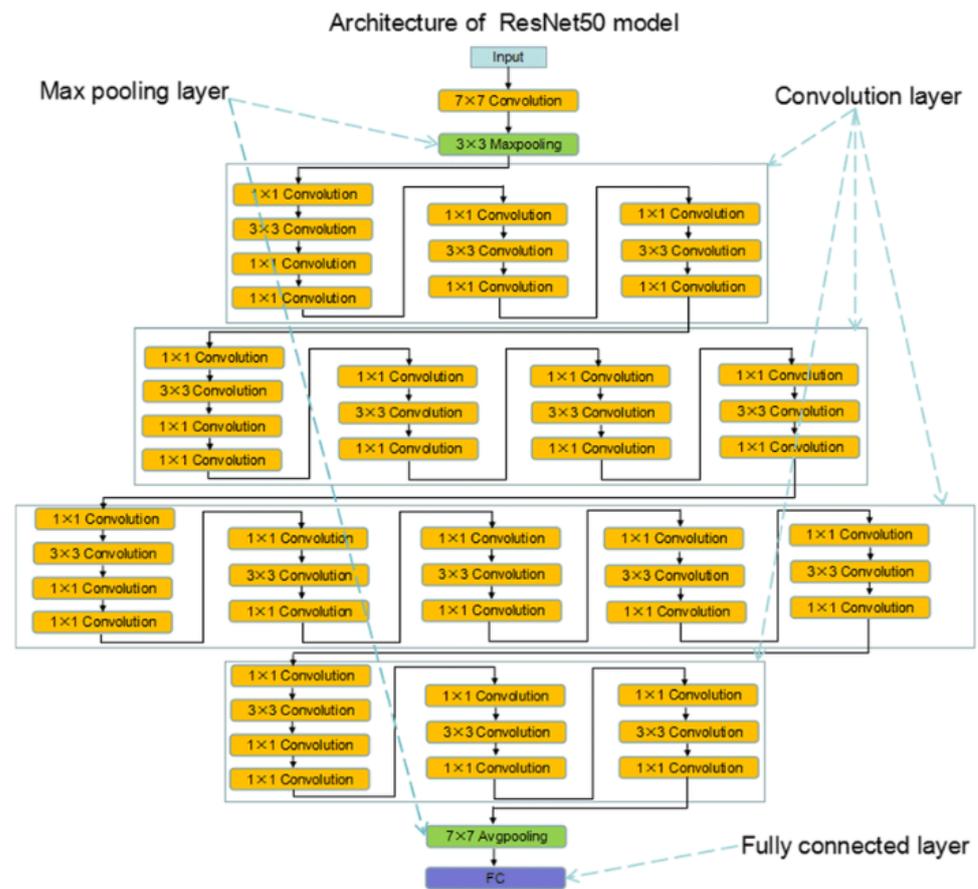
**Figure 10.** ResNet50 architecture [53].

The model gave an accuracy of approximately 74% on the tested images, displaying errors presented in the last image where an umbrella was classified as a banana with an accuracy of over 40% (see Figure 11). This error nevertheless occurs because the dataset on which the model was trained contains only three types of classes (apple, orange, and banana), and the model tries to classify all objects into one of the categories from antecedent experience. The existence of another potential class termed "non-fruit" would have helped the model render a more accurate classification of objects.
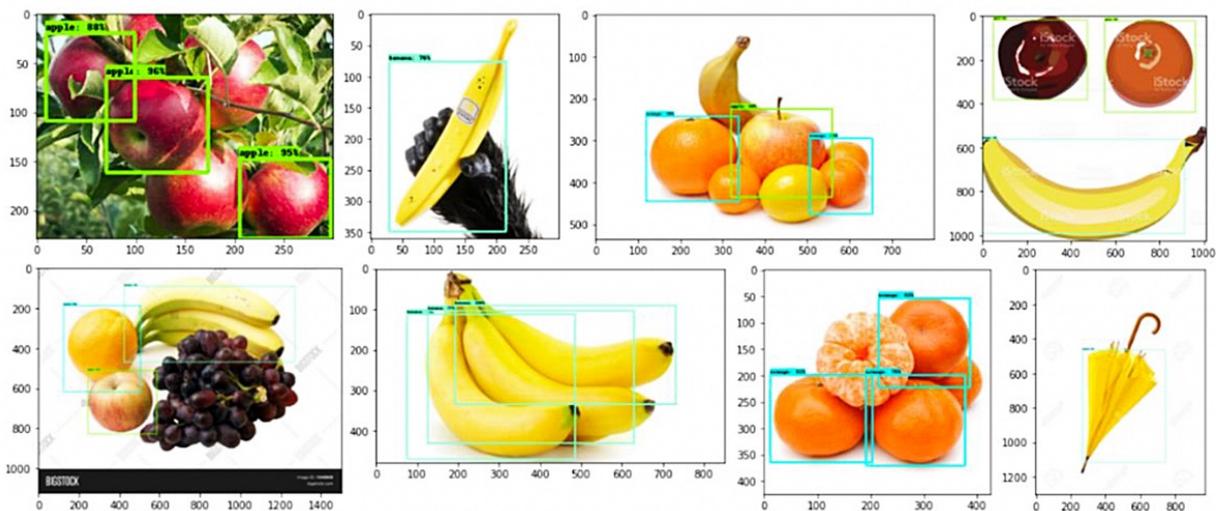


**Figure 11.** Image detection result in images [54].

For real-time object detection in the application, the faster R-CNN ResNet50 V1 640 × 640 model was used. This is a model based on the same ResNet50 architecture (Figure 10), but its algorithm is faster R-CNN. The model was trained on the COCO 2017 dataset, being able to detect up to 90 object classes.

Algorithms like R-CNNs (region-based convolutional neural networks) are recognized for their performance in object detection accuracy, but the inference process was slow. Faster R-CNN is a much faster algorithm compared with its predecessors (R-CNN and fast R-CNN). Indeed, unlike them, it does not use the selective search technique (selective search consumes a lot of time and computational power). That technique involves extracting thousands of regions from an image and making predictions by a convolutional neural network (CNN) on each region. Therefore, it is understandable why the selective search technique is so computationally expensive [55]. The improvement brought with the faster R-CNN algorithm is the replacement of the selective search with a region proposal network (RPN) [56].

This network performs a binary classification (which identifies whether a certain region contains an object or not) on a series of regions called anchor boxes, which have different sizes. Following the classification performed by this network, only regions that have been classified as enclosing objects (Figure 12) are sent to the region of interest (ROI) pooling layer [57].
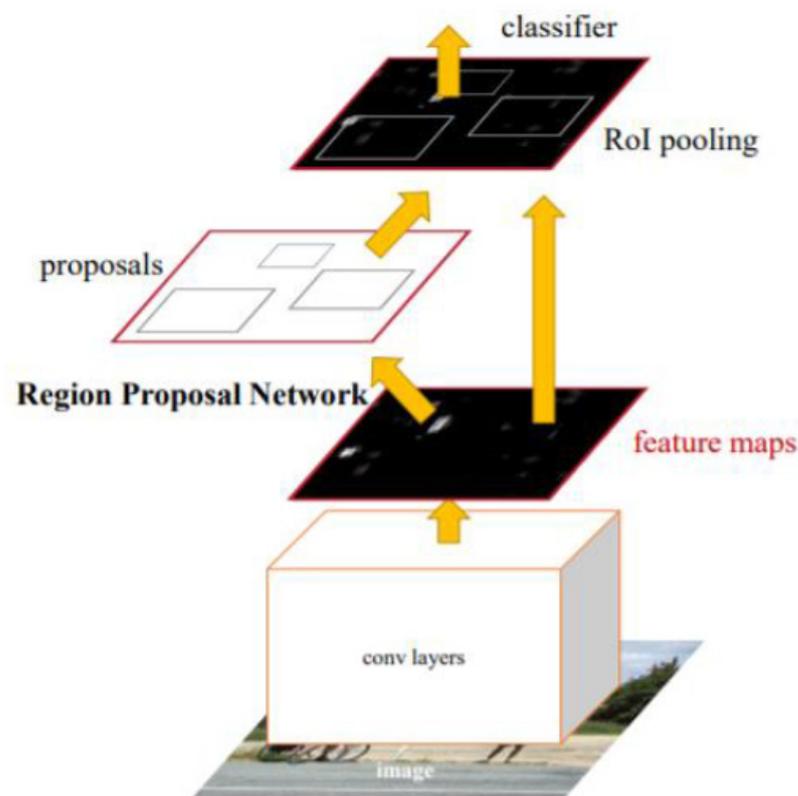


**Figure 12.** Faster R-CNN architecture [58].

Figure 13 shows object detections made with the pre-trained faster R-CNN ResNet50 V1 640 × 640 model, in real-time provided images, displaying good very good accuracy for real-life scenarios (above 80%).
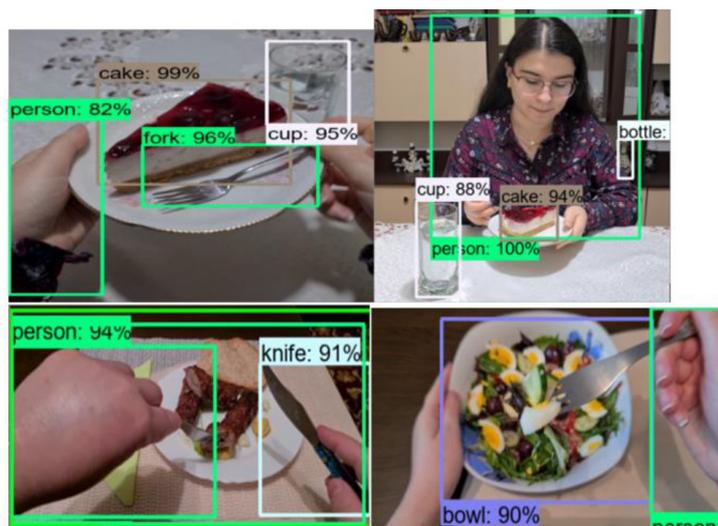
**Figure 13.** Object detection in images.

### 4.3. Recognition of Human Activity in Video

Lastly, one of the studies on AI endeavors aimed to offer a development approach for allowing the system to be able to recognize important human activities in real time, like eating, an activity useful in an AD context (for both patients and their companions). The research assumption was that it is possible to develop such a system with predictive capabilities above 70%.

Recognition of activity (events) is a branch of computer vision and involves the recognition of actions or events in a video depending on the context [59]. To recognize the activity in the video, the model was trained on the dataset "Kinetics 400", which, as the name suggests, contains 400 human activities with at least 400 videos each, recorded in videos on YouTube [60]. Of the 400 activities, only the activities related to eating or drinking, such as "eating burger" and "drinking", were considered for the current research. The problem of classifying human actions in video is complex—it involves the analysis of a video containing both a spatial component (frames) and a temporal component (the order of the frames) [61]. The first tested model was based on the ResNet34 network architecture, which is made up of 34 convolutional layers (Figure 14).

| Layer name | Resnet-34 | | |
|---|---|---|---|
| conv1 | 7x7, 64, stride 2 | | |
| pool1 | 3x3, max pool, stride 2 | | |
| conv2_x | $3 \times 3, 64$ | $\times 3$ | |
| | $3 \times 3, 64$ | | |
| conv3_x | $3 \times 3, 128$ | $\times 4$ | |
| | $3 \times 3, 128$ | | |
| conv4_x | $3 \times 3, 256$ | $\times 6$ | |
| | $3 \times 3, 256$ | | |
| conv5_x | $3 \times 3, 512$ | $\times 3$ | |
| | $3 \times 3, 512$ | | |
| fc1 | 4x1, 512, stride 1 | | |
| pool time | 1x10, avg pool, stride 1 | | |
| fc2 | 1x1, 50 | | |

**Figure 14.** ResNet34 architecture [62].

The convolutional kernel of 2D convolutional networks (used for image processing) is replaced by a 3D kernel. Researchers have found that, while deep network architectures used for image classification (such as ResNet) provide high accuracy when trained on very large datasets (such as ImageNet for images), a large enough video dataset is suitable for performing video classification [63,64]. This model presented in [63] obtained an accuracy of 78.4% on the test set of the Kinetics 400 dataset [63]. A comparison was made between 2D neural networks pretrained on the ImageNet dataset for image classification and 3D neural networks pretrained on Kinetics 400 for video classification, proving that classification of human actions in real-time videos can be performed with a very good performance (see Figure 15).
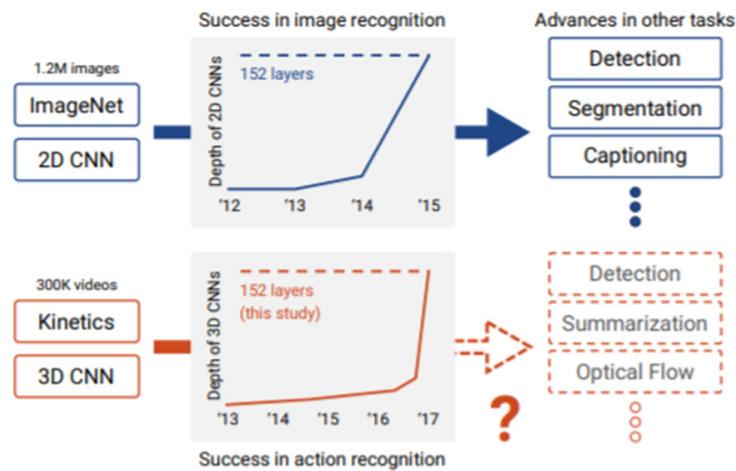


**Figure 15.** Comparison made in the study of the training of the model of classification of human actions in videos [63].

Figure 16 presents a real-life example where the contextual activity is recognized as "tasting food", being correctly interpreted as a general activity "eat".
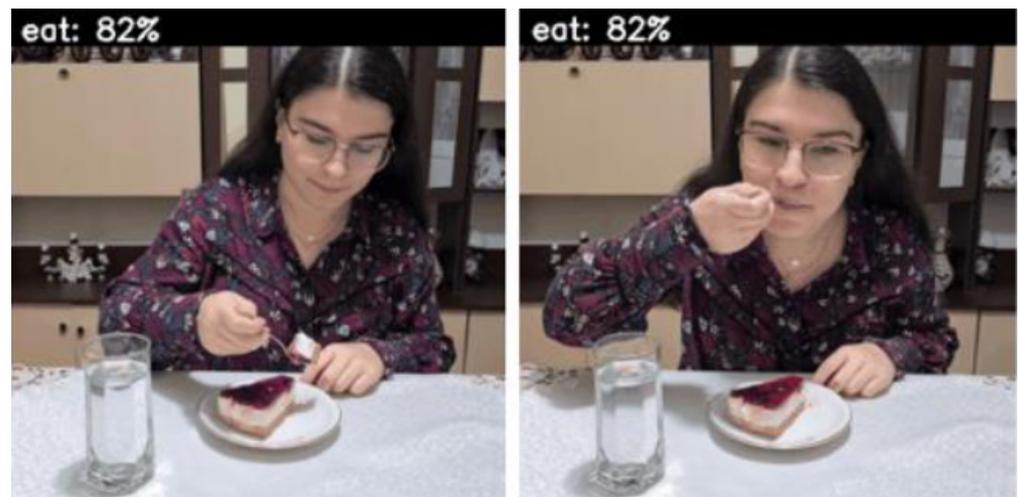


**Figure 16.** Frames from a video after the recognition of human activity.

The second model tested for video human activity recognition was developed in [65], where a new two-stream inflated 3D ConvNet architecture based on the inflation of 2D convolutional networks is introduced, in which the 2D core of deep networks for image classification was expanded to a 3D one, making it possible to process the spatial and temporal component of videos (see Figure 17) [65].
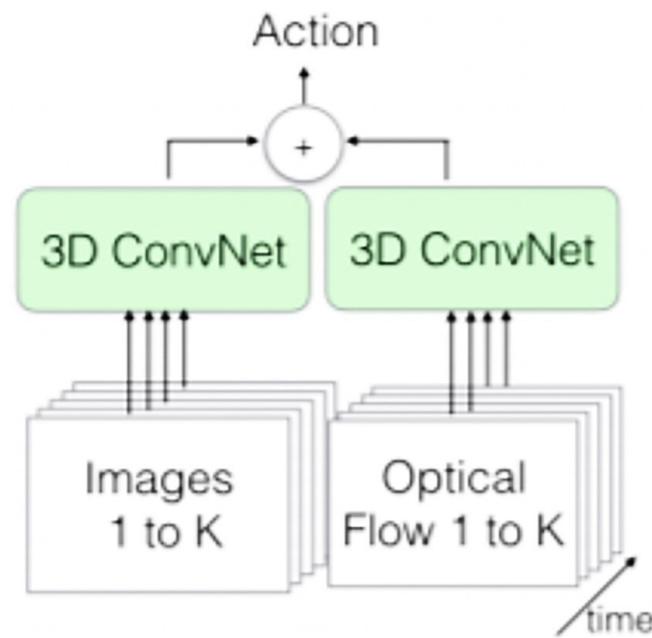
**Figure 17.** Two-stream I3D architecture [65].

This model achieved an accuracy of 74.2% on the Kinetics 400 test set [65]. Still, by comparing the two models of human activity recognition in video, the first model based on 3D spatial-temporal convolutional neural networks recorded both a higher accuracy on the Kinetics 400 test set and a higher computational speed, being faster than the second model.

### 4.4. System Architecture

The image classification and the detection of human activity deep learning algorithms presented above can be successfully embedded into a real-time intelligent system, designed for assisting people with Alzheimer's disease (Figure 18).
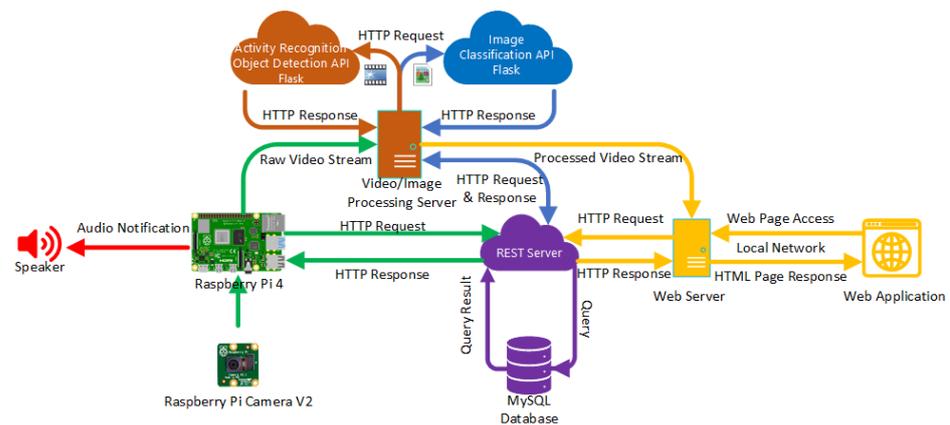


**Figure 18.** System diagram.

Using the architecture presented above as a starting point, we developed an intelligent mobile/portable system designed specifically to assist persons developing Alzheimer's disease symptoms. As presented in Figure 18, the system is comprised of several main components: (a) the deep learning models exposed through a Python Flask interface, (b) the video surveillance system interacting with the deep learning models and with the database, (c) the system database exposed also through REST services, and (d) the web application for both patients and caretakers.

To create the video surveillance and audio notification device, a different type of equipment was used (Table 1).

**Table 1.** Video surveillance and notification—portable device equipment.

| Equipment | Characteristics |
| --- | --- |
| Raspberry Pi 4 | 64-bit quad-core processor, dual-display support at resolutions up to 4 K via a pair of micro-HDMI ports, hardware video decode at up to 4 Kp60, up to 8 GB of RAM, dual-band 2.4/5.0 GHz wireless LAN, Bluetooth 5.0, Gigabit Ethernet, USB 3.0 and PoE capability (via a separate PoE HAT add-on) |
| Pi Camera | Raspberry Pi Camera Board V2 features an ultra-high quality 8-megapixel Sony IMX219 image sensor (up from 5 MP on the V1 camera board) and a fixed focus camera lens. The V2 camera module is capable of 3280 × 2464 pixel static images, and supports 1080 p30, 720 p60, and 640 × 480 p90 video. |
| Portable speaker | wireless portable speaker BT25G/00 with anti-clipping function for loud, distortion-free sounds and built-in rechargeable battery |
| External Battery | 10,050 mAh external battery, length: 10 cm, width: 7 cm, height: 4 cm, weight: 250 g |

*4.5. The Monitoring Application*

As presented in Figure 18, to access the platform functionality, a web application was developed using specific technologies like REST services, Spring Boot, Spring Data JPA, Apache Maven, Thymeleaf, HTML, MySql, Flask, WebSockets, ImageZMQ, OpenCV, Google Text to Speech, and OMXPlayer.

When accessing the application, a web page opens where the caretaker can monitor the Alzheimer's disease patient in real time. On the "surveillance" page, two video streams are displayed, the first one having the image processed following the recognition of human activity, by noting the action of interest performed if it occurs (the patient eats or drinks), and the probability score of the prediction made by the model for video activity recognition. The second video stream consists of the same images, but after processing for the detection of objects in the images and the binary classification of the images (food or not-food). Surveillance images can be displayed from two perspectives depending on how the surveillance device is positioned: (a) a fixed position (see Figure 19), which makes possible to recognize the human activity in the provided video stream or to perform binary image classification (Figures 20 and 21); (b) mobile positioning, where the device is mobile and is filmed from the perspective of the monitored person because he/she is wearing it permanently (Figures 22–24).
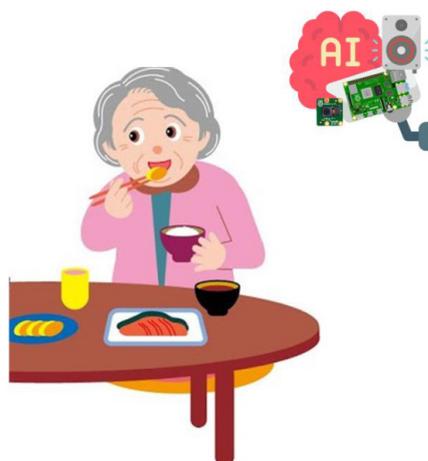


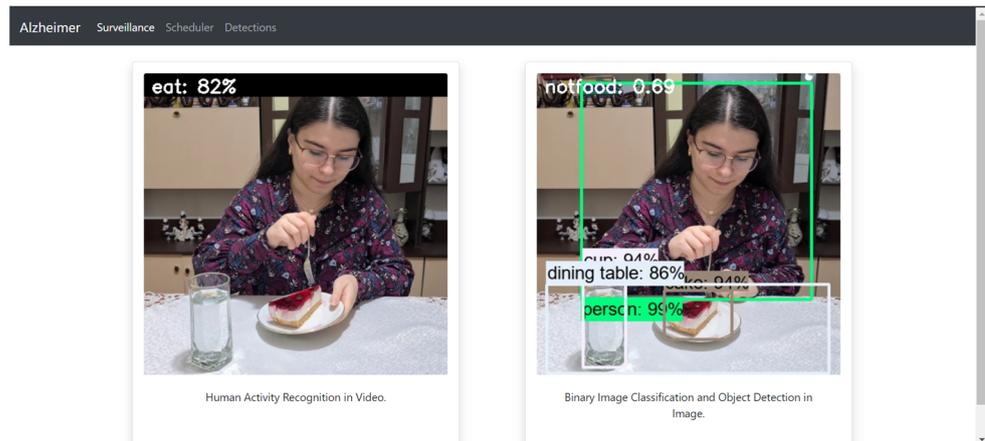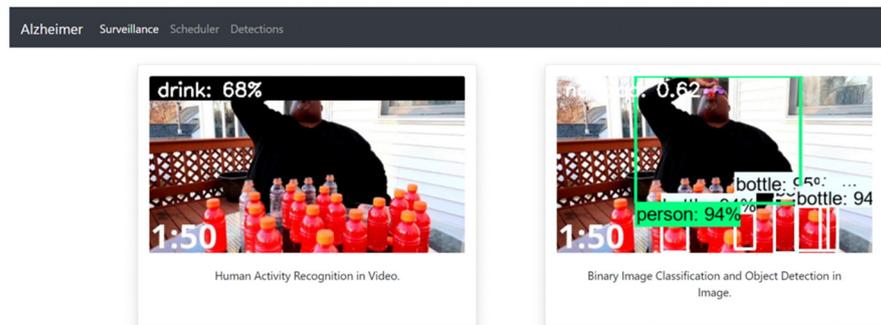**Figure 19.** Fixed position monitoring device.

**Figure 20.** Surveillance 1.



**Figure 21.** Surveillance 2.



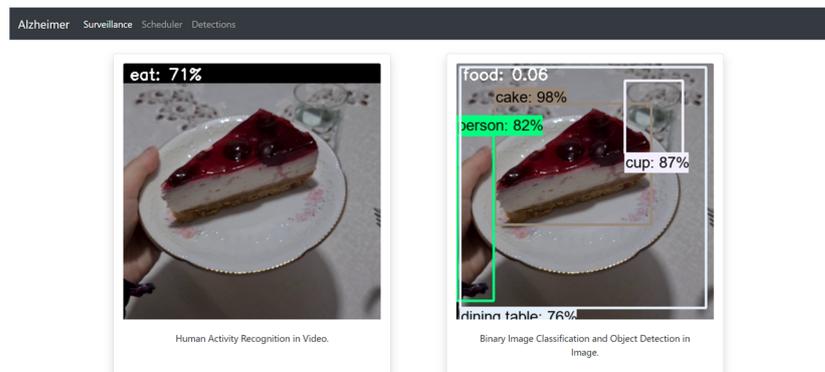**Figure 22.** Mobile surveillance device.
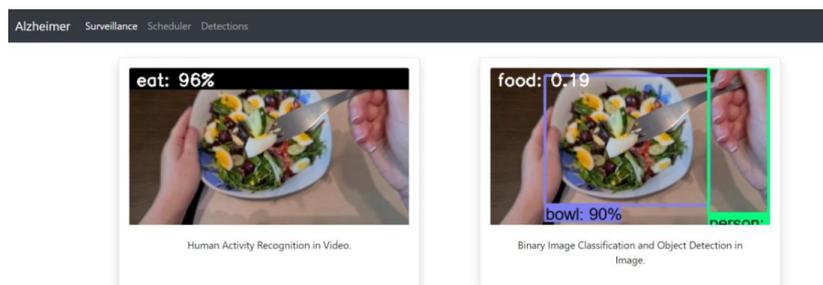


**Figure 23.** Surveillance 3.

**Figure 24.** Surveillance 4.

By accessing the "scheduler" page from the main menu, the caregiver can modify the settings made for the established rules related to the distance between two consecutive meals and the duration of a single meal, in order to avoid underfeeding or overfeeding the patient, and how often the patient hydrates to avoid dehydration (Figure 25). Through the scheduler page, the caregiver has the possibility to modify the following aspects:

- The email address through which he receives notifications related to the patient's activity;
- The period, in seconds, during which the patient's activity is scanned. The value can be between 1 and 120 s;
- The time at which the Alzheimer's disease patient usually starts his sleep schedule to avoid sending audio notifications that can wake him up during sleep, such as notifications that he has not eaten or drunk for a long time;
- The time at which the Alzheimer's disease patient usually wakes up;
- The minimum number of hours between two consecutive meals that the patient can have to avoid overeating. The value can be between 1 and 10 h;
- The maximum number of hours between two consecutive meals that the patient can have to avoid malnutrition. The value can be between 2 and 10 h;
- The period in minutes in which the patient can eat at one meal to avoid overeating. The value can be between 1 and 60 min;
- The number of minutes between the moments in which the patient drinks hydrating liquids to avoid dehydration if the set period is exceeded without him drinking anything. The value can be between 1 and 120 min;
- The number of minutes between the notifications informing him that he has to eat to avoid malnutrition (the number of minutes between notifications ensures that the patient has time to prepare his meal from the first notification to the next to avoid stressing him). The value can be between 1 and 30 min;
- The number of minutes between notifications informing him that he needs to hydrate to avoid dehydration (the number of minutes between notifications ensures that the patient has time to prepare his drinking liquids from the first notification to the next to avoid stressing him). The value can be between 1 and 15 min;
- Modification of the audio message that informs the patient that he must stop eating to avoid overeating. This is preset with the value "You should stop eating", but the caretaker can customize the message in a suggestive way for the person for whom they care;
- Modification of the message received at the email address for the caregiver that informs him that the patient must stop eating to avoid overfeeding. This is preset with the value "The patient is eating too often", but it can be personalized in a suggestive way according to the caregiver's preferences;
- Modification of the message received at the email address for the caregiver that informs him that the patient forgot to eat to avoid sublimation. This is preset with the value "The patient forgot to eat", but it can be personalized in a suggestive way according to the caregiver's preferences;
- Modification of the message received at the email address for the caregiver notifying him that the patient forgot to drink hydrating liquids to avoid dehydration. This is

preset with the value "The patient forgot to drink", but it can be personalized in a suggestive way according to the caregiver's preferences;

- The number of hours after which data related to videos and notifications saved in the database are deleted from the time of recording, with the value being between 6 and 72 h.
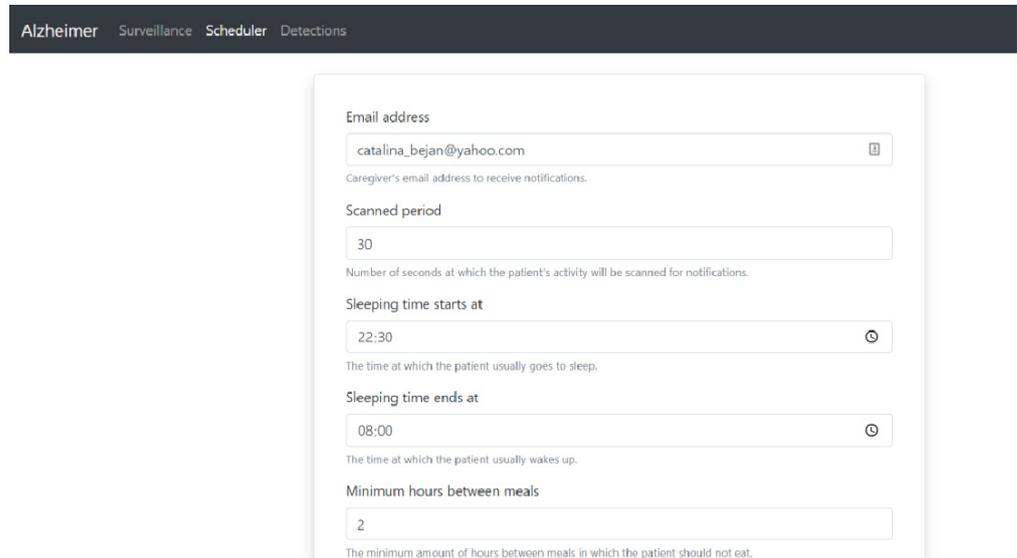


**Figure 25.** Scheduler interface.

The detections page is accessed from the main menu and includes the recordings of moments of interest (when the person eats or drinks) together with the time they were recorded, the action identified by the human activity recognition model, and the score of the prediction made. Up to five videos are displayed on each page (with pagination) in a table as shown in Figure 26.
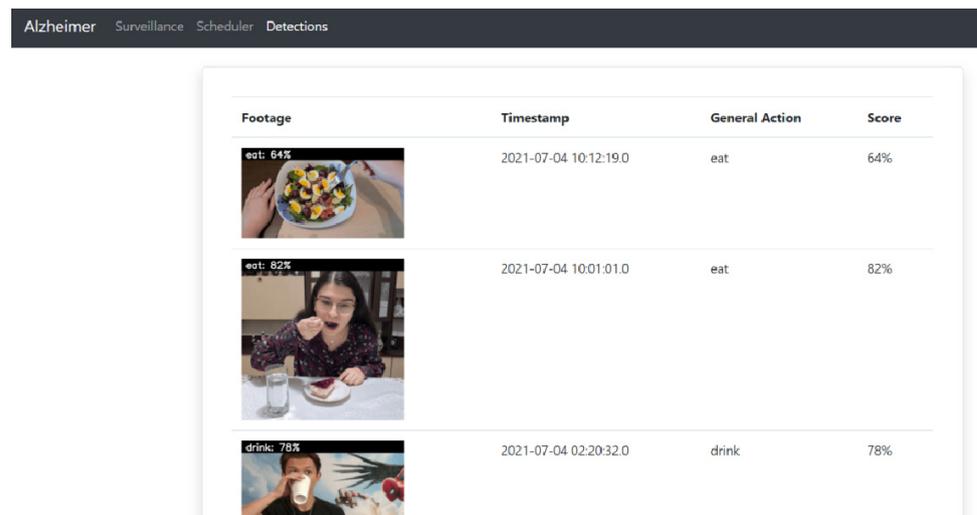


**Figure 26.** Detections.

The application can be accessed from any device with access to an Internet browser because it is scalable and responsive depending on the device on which it is displayed, as shown in Figures 27–29, where three frames from the surveillance page are displayed.
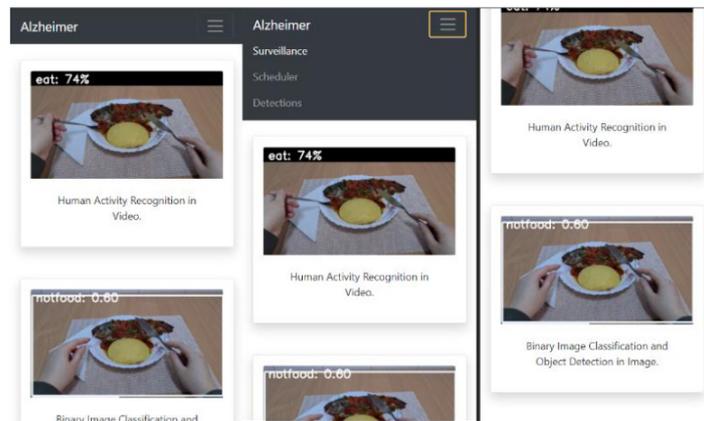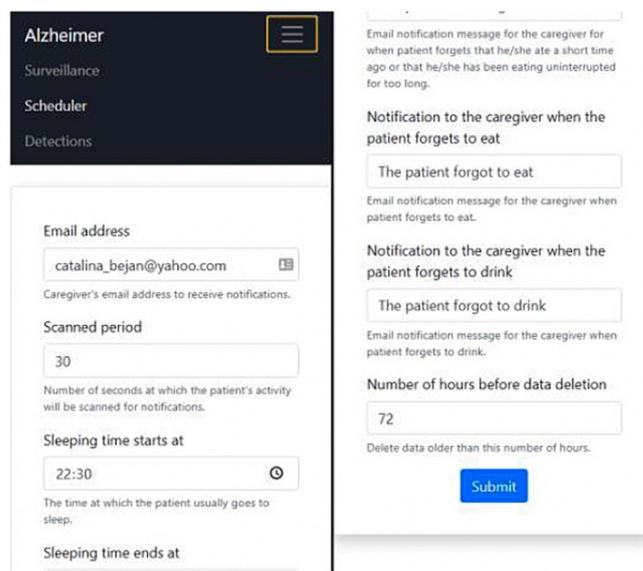
**Figure 27.** Responsive surveillance.
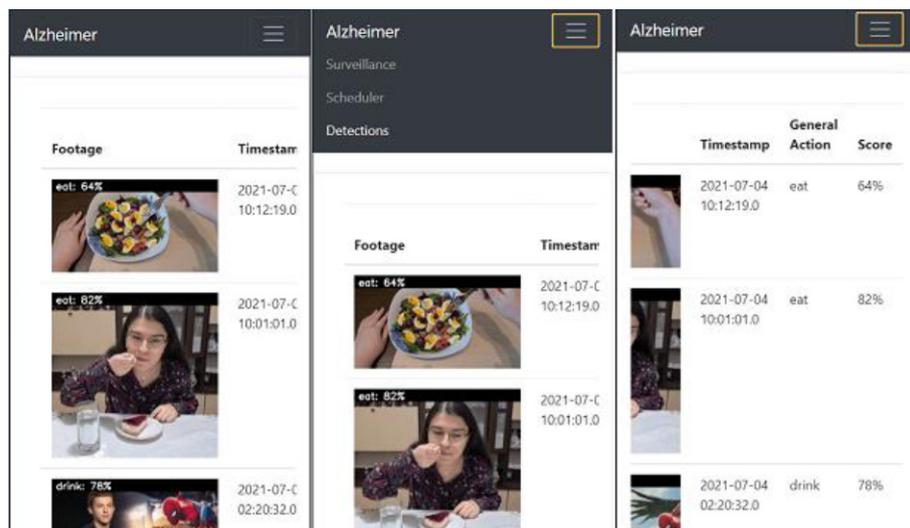


**Figure 28.** Responsive scheduler.



**Figure 29.** Responsive detections.

Figure 30 shows the types of notifications the caregiver can receive, for which the subject is set in the database and cannot be changed by the caregiver.
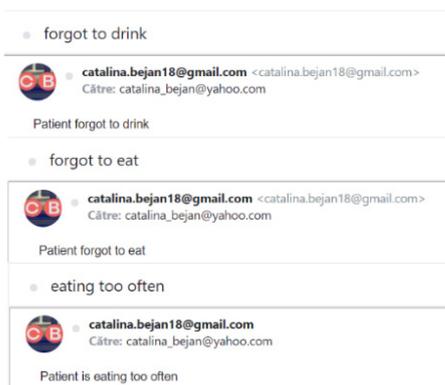
**Figure 30.** Email notifications.

The person with Alzheimer's disease who is monitored through this system will receive audio notifications from the monitoring device, which has an audio speaker (speaker and headphones) connected. The notifications received remind them of when they should eat and hydrate to ensure a healthy nutritional style:

1.  Forgot to eat notification—This type of notification is received when the person with Alzheimer's disease is not sleeping and has not eaten for longer than the maximum number of hours between meals by default or set by the caregiver in the application. The notification is of the form "You should eat" by default, but can take any personalized but suggestive form provided by the caregiver. The notification is repeated if the patient continues not to eat after several minutes set in the web application between reminder notifications to eat.

2.  Forgotten hydration notification—This type of notification is received when the person with Alzheimer's disease is not sleeping and has not had a drink for longer than the default number of minutes between meals or set by the caregiver in the app. The notification is of the form "You should drink" by default, but can take any personalized but suggestive form provided by the caregiver. The notification is repeated if the patient continues not to drink after several minutes set in the web application between drinking reminder notifications.

3.  The notification that he forgot that he ate a while ago or that he forgot that he ate for too long—This type of notification is received even when the person is in the sleep schedule (the interval set by the caregiver in the web application), because if it is identified that they eat too often or too much, then there is no risk of waking the person from sleep. The notification is sent if the person is recorded as eating the current meal in a shorter time than the set minimum number of hours between two consecutive meals, so they are eating too often or having too many meals per day. At the same time, this notification also appears if the time spent eating at one meal exceeds the number of normal minutes for eating. This indicates that the person is eating too much, and the notification appears repeatedly with an interval equivalent to the number of seconds between checks on the database set by the caregiver in the app. The message received in this case is "You should stop eating" by default or any other suggestive message set by the caregiver in the web application.

## 5. Discussion and Conclusions

Neural network training is a lengthy process, no matter the purpose of the network—image classification, object detection, or activity analysis. Depending on the machine hardware, it can take up to several days. After training the networks, the resulting model can be deployed with very fast response times (two seconds).

After performing the checks on the performance score, the confusion matrix, the viewing of the misclassified images, and the prediction on an image that is not found in the dataset, certain aspects can be concluded. The binary classification model whose

architecture is based on the architecture of the VGG16 model and that has knowledge transfer from the same model is one with high accuracy and low error rate—aspects that make it a model suitable for use in real life.

Object detection in images is a method of great interest for the requirements of the times in terms of artificial vision, being a complex problem. Using transfer learning methodology and high-performance deep learning networks, it can be concluded that the models for object detection have a high performance, with the objects being detected in almost real time. Recognition of human activity in a video is a complex issue in the field of artificial vision. This involves both image processing and the time component of a video (image order). These models require a high computational power for training, but also for testing if a near-real-time response speed is desired. The technological platform, based on deep learning AI models, implemented in this paper aims to help people with Alzheimer's disease regain their independence, but also their caregivers by easing their duties. Thus, the system allows video surveillance of the Alzheimer's disease patient and, by means of artificial intelligence, the moments in which the supervised person is fed or hydrated are identified.

The system is designed to prevent malnutrition, overeating, or dehydration of a person with Alzheimer's disease by notifying them with an audible message and reminding them of the actions they need to take for a healthy lifestyle. The need to develop such a system is closely linked to the fact that the number of people with Alzheimer's disease is constantly growing and there is currently no surveillance system that recognizes their actions and helps them live independently. The application is distinguished by the new concept it introduces, that is, the recognition of human activity to help people with Alzheimer's or dementia and their caregivers. The system is designed to recognize the activity from several perspectives; that is, the one in which the surveillance device is in a fixed position and the one in which the device is mobile and is carried by the patient. Such a system can be improved by creating a large enough dataset specific to the actions of interest for training the models. A noteworthy future implementation (and currently a limitation) is the lack of a proper identification mechanism like biometric identification, such as fingerprint, face, voice, or iris (therefore, avoiding the imminent risk of the patient forgetting his authentication credentials). The main purpose of connecting the patient to the application will be to remember the moment when he/she last ate or drank, through a video recording that, for the time being, can only be accessed by the caregiver. It is also desired to introduce a doctor role in the application that can medically manage the patient's health. The introduction of statistics of recorded activities will be very helpful to the doctor for the eventual establishment of a diet according to the patient's habits. Thus, the system should allow the possibility of introducing a pre-established diet by identifying the foods they eat (from the results of object detection and image classification algorithms).

As a technical limitation, the weight of the portable device, mainly induced by the battery weight (approximately 250 g), and its size (10 cm in length, 7 cm width, and 4 cm height) are also to be mentioned.

Regarding future enhancements, the weight and dimensions of the portable assistant device should be lowered through the use of smaller batteries and development boards (for example, Raspberry Pi Zero). In addition, the device could be prototyped for use with modern smartphones' cameras and processing power. Simple sensors can be added to the overall architecture for detecting when different doors are opened and the software application, besides warning/attention messages, could stream video content to the user in the eating/drinking moments.

**Author Contributions:** Conceptualization, D.M., C.B., N.M., S.-M.P. and D.C.; Data curation, D.M., C.B., S.-M.P. and D.C.; Formal analysis, D.M. and C.B.; Funding acquisition, C.Z.; Methodology, C.B., N.M. and D.C.; Project administration, D.M.; Resources, D.C.; Software, D.M., C.B. and D.C.; Supervision, D.C.; Validation, C.Z. and S.-M.P.; Visualization, N.M. and C.Z.; Writing—original draft, C.B.; Writing—review and editing, S.-M.P., M.V. and D.C. All authors have read and agreed to the published version of the manuscript.

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

## References

1. Ulep, M.G.; Saraon, S.K.; McLea, S. Alzheimer Disease. *J. Nurse Pract.* **2018**, *14*, 129–135. [CrossRef]
2. Dementia Statistics | Alzheimer's Disease International (ADI). Available online: https://www.alzint.org/about/dementia-facts-figures/dementia-statistics/ (accessed on 30 July 2022).
3. Biggs, S.; Carr, A.; Haapala, I. Dementia as a Source of Social Disadvantage and Exclusion. *Australas. J. Ageing* **2019**, *38*, 26–33. [CrossRef] [PubMed]
4. Vally, M.; Kathrada, F. Understanding Alzheimer Disease. *S. Afr. Fam. Pract.* **2019**, *61*, 28–34. [CrossRef]
5. What You Need to Know about Dementia and Eating Disorders—The Alzheimer's Site News. Available online: https://blog.thealzheimerssite.greatergood.com/eating-disorders/ (accessed on 28 July 2022).
6. Wilson, K.; Dewing, J. Strategies to Prevent Dehydration in Older People with Dementia: A Literature Review. *Nurs. Older People* **2019**. Erratum in *Nurs. Older People* **2020**, *32*, 27–33. [CrossRef]
7. Aljojo, N.; Alotaibi, R.; Alharbi, B.; Alshutayri, A.; Tariq Jamal, A.; Banjar, A.; Khayyat, M.; Zainol, A.; Al-Roqy, A.; Al-Magrabi, R.; et al. Alzheimer Assistant: A Mobile Application Using Machine Learning. *Rev. Română Inform. Autom.* **2020**, *30*, 7–26. [CrossRef]
8. Dementia: Technology to Stay Independent—SCIE. Available online: https://www.scie.org.uk/dementia/support/technology/stay-independent (accessed on 28 July 2022).
9. Sharma, S.; Dudeja, R.K.; Aujla, G.S.; Bali, R.S.; Kumar, N. DeTrAs: Deep Learning-Based Healthcare Framework for IoT-Based Assistance of Alzheimer Patients. *Neural Comput. Appl.* **2020**. [CrossRef]
10. Roopaei, M.; Rad, P.; Prevost, J.J. A Wearable IoT with Complex Artificial Perception Embedding for Alzheimer Patients. In Proceedings of the 2018 World Automation Congress (WAC), Stevenson, WA, USA, 3–6 June 2018; IEEE Computer Society: Washington, DC, USA, 2018; Volume 2018, pp. 28–33.
11. Gollapudi, S. *Learn Computer Vision Using OpenCV: With Deep Learning CNNs and RNNs*; Apress: New York, NY, USA, 2019.
12. McAfee, R.P. Review of AI Superpowers: China, Silicon Valley and the New World Order, by Kai-Fu Lee. *Bus. Econ.* **2019**, *54*, 185–190. [CrossRef]
13. GitHub. Rileykwok/Food-Classification: A CNN Model Is Trained to Classify 3 Classes of Food in the Food-101 Data Set. Available online: https://github.com/rileykwok/Food-Classification (accessed on 28 July 2022).
14. Vinny, P.W.; Vishnu, V.Y.; Padma Srivastava, M.V. Artificial Intelligence Shaping the Future of Neurology Practice. *Med. J. Armed Forces India* **2021**, *77*, 276–282. [CrossRef] [PubMed]
15. Fabrizio, C.; Termine, A.; Caltagirone, C.; Sancesario, G. Artificial Intelligence for Alzheimer's Disease: Promise or Challenge? *Diagnostics* **2021**, *11*, 1473. [CrossRef] [PubMed]
16. Sharma, D.K.; Chatterjee, M.; Kaur, G.; Vavilala, S. Deep Learning Applications for Disease Diagnosis. In *Deep Learning for Medical Applications with Unique Data*; Elsevier: Amsterdam, The Netherlands, 2022.
17. Vyas, A.H.; Mehta, M.A.; Kotecha, K.; Pandya, S.; Alazab, M.; Gadekallu, T.R. Tear Film Breakup Time-Based Dry Eye Disease Detection Using Convolutional Neural Network. *Neural Comput. Appl.* **2022**, *2022*, 1–19. [CrossRef]
18. de Fauw, J.; Ledsam, J.R.; Romera-Paredes, B.; Nikolov, S.; Tomasev, N.; Blackwell, S.; Askham, H.; Glorot, X.; O'Donoghue, B.; Visentin, D.; et al. Clinically Applicable Deep Learning for Diagnosis and Referral in Retinal Disease. *Nat. Med.* **2018**, *24*, 1342–1350. [CrossRef]
19. Storås, A.M.; Strümke, I.; Riegler, M.A.; Grauslund, J.; Hammer, H.L.; Yazidi, A.; Halvorsen, P.; Gundersen, K.G.; Utheim, T.P.; Jackson, C.J. Artificial Intelligence in Dry Eye Disease. *Ocul. Surf.* **2022**, *23*, 74–86. [CrossRef] [PubMed]

20. le Berre, C.; Sandborn, W.J.; Aridhi, S.; Devignes, M.D.; Fournier, L.; Smaïl-Tabbone, M.; Danese, S.; Peyrin-Biroulet, L. Application of Artificial Intelligence to Gastroenterology and Hepatology. *Gastroenterology* **2020**, *158*, 76–94.e2. [CrossRef]

21. Thrall, J.H.; Li, X.; Li, Q.; Cruz, C.; Do, S.; Dreyer, K.; Brink, J. Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success. *J. Am. Coll. Radiol.* **2018**, *15*, 504–508. [CrossRef]

22. Yeh, C.H.; Yu, S.X.; Lin, M.C. Meibography Phenotyping and Classification From Unsupervised Discriminative Feature Learning. *Transl. Vis. Sci. Technol.* **2021**, *10*, 4. [CrossRef] [PubMed]

23. Ramana, K.; Kumar, M.R.; Sreenivasulu, K.; Gadekallu, T.R.; Bhatia, S.; Agarwal, P.; Idrees, S.M. Early Prediction of Lung Cancers Using Deep Saliency Capsule and Pre-Trained Deep Learning Frameworks. *Front. Oncol.* **2022**, *12*, 2641. [CrossRef]

24. Jothi, G.; Inbarani, H.H. Soft Set Based Feature Selection Approach for Lung Cancer Images. *arXiv* **2012**, arXiv:1212.5391.

25. Kaggle. Food Images (Food-101). Available online: https://www.kaggle.com/datasets/kmader/food41 (accessed on 28 July 2022).

26. Kaggle. Furniture Detector. Available online: https://www.kaggle.com/datasets/akkithetechie/furniture-detector (accessed on 28 July 2022).

27. IMDB-WIKI—500k+ Face Images with Age and Gender Labels. Available online: https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/ (accessed on 28 July 2022).

28. Rothe, R.; Timofte, R.; van Gool, L. Deep Expectation of Real and Apparent Age from a Single Image without Facial Landmarks. *Int. J. Comput. Vis.* **2018**, *126*, 144–157. [CrossRef]

29. Rothe, R.; Timofte, R.; van Gool, L. DEX: Deep EXpectation of Apparent Age from a Single Image. In Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), Santiago, Chile, 7–13 December 2015; Volume 2015.

30. Böhle, M.; Eitel, F.; Weygandt, M.; Ritter, K. Layer-Wise Relevance Propagation for Explaining Deep Neural Network Decisions in MRI-Based Alzheimer's Disease Classification. *Front. Aging Neurosci.* **2019**, *10*, 194. [CrossRef]

31. Nguyen, M.; He, T.; An, L.; Alexander, D.C.; Feng, J.; Yeo, B.T.T. Predicting Alzheimer's Disease Progression Using Deep Recurrent Neural Networks. *Neuroimage* **2020**, *222*, 117203. [CrossRef] [PubMed]

32. Shi, B.; Chen, Y.; Zhang, P.; Smith, C.D.; Liu, J. Nonlinear Feature Transformation and Deep Fusion for Alzheimer's Disease Staging Analysis. *Pattern Recognit.* **2017**, *63*, 487–498. [CrossRef]

33. Mufidah, R.; Wasito, I.; Hanifah, N.; Faturrahman, M. Structural MRI Classification for Alzheimer's Disease Detection Using Deep Belief Network. In Proceedings of the 11th International Conference on Information and Communication Technology and System, ICTS 2017, Surabaya, Indonesia, 31 October 2017; Volume 2018.

34. Nawaz, H.; Maqsood, M.; Afzal, S.; Aadil, F.; Mehmood, I.; Rho, S. A Deep Feature-Based Real-Time System for Alzheimer Disease Stage Detection. *Multimed. Tools Appl.* **2021**, *80*, 35789–35807. [CrossRef]

35. Liang, S.; Gu, Y. Computer-Aided Diagnosis of Alzheimer's Disease through Weak Supervision Deep Learning Framework with Attention Mechanism. *Sensors* **2021**, *21*, 220. [CrossRef]

36. Katabathula, S.; Wang, Q.; Xu, R. Predict Alzheimer's Disease Using Hippocampus MRI Data: A Lightweight 3D Deep Convolutional Network Model with Visual and Global Shape Representations. *Alzheimer's Res. Ther.* **2021**, *13*, 104. [CrossRef] [PubMed]

37. Gao, S.; Lima, D. A Review of the Application of Deep Learning in the Detection of Alzheimer's Disease. *Int. J. Cogn. Comput. Eng.* **2022**, *3*, 1–8. [CrossRef]

38. Saleem, T.J.; Zahra, S.R.; Wu, F.; Alwakeel, A.; Alwakeel, M.; Jeribi, F.; Hijji, M. Deep Learning-Based Diagnosis of Alzheimer's Disease. *J. Pers. Med.* **2022**, *12*, 815. [CrossRef]

39. Chien, Y.W.; Hong, S.Y.; Cheah, W.T.; Yao, L.H.; Chang, Y.L.; Fu, L.C. An Automatic Assessment System for Alzheimer's Disease Based on Speech Using Feature Sequence Generator and Recurrent Neural Network. *Sci. Rep.* **2019**, *9*, 19597. [CrossRef]

40. Roshanzamir, A.; Aghajan, H.; Soleymani Baghshah, M. Transformer-Based Deep Neural Network Language Models for Alzheimer's Disease Risk Assessment from Targeted Speech. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 92. [CrossRef]

41. Allegra, D.; Battiato, S.; Ortis, A.; Urso, S.; Polosa, R. A Review on Food Recognition Technology for Health Applications. *Health Psychol. Res.* **2020**, *8*, 3. [CrossRef]

42. Munkevik, P.; Hall, G.; Duckett, T. A Computer Vision System for Appearance-Based Descriptive Sensory Evaluation of Meals. *J. Food Eng.* **2007**, *78*, 246–256. [CrossRef]

43. Kong, F.; Tan, J. DietCam: Automatic Dietary Assessment with Mobile Camera Phones. *Pervasive Mob. Comput.* **2012**, *8*, 147–163. [CrossRef]

44. Lu, Y.; Stathopoulou, T.; Vasiloglou, M.F.; Christodoulidis, S.; Blum, B.; Walser, T.; Meier, V.; Stanga, Z.; Mougiakakou, S.G. An Artificial Intelligence-Based System for Nutrient Intake Assessment of Hospitalised Patients. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, Berlin, Germany, 23–27 July 2019.

45. Jo, T.; Nho, K.; Saykin, A.J. Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. *Front. Aging Neurosci.* **2019**, *11*, 220. [CrossRef] [PubMed]

46. Jack, C.R.; Bennett, D.A.; Blennow, K.; Carrillo, M.C.; Dunn, B.; Haeberlein, S.B.; Holtzman, D.M.; Jagust, W.; Jessen, F.; Karlawish, J.; et al. NIA-AA Research Framework: Toward a Biological Definition of Alzheimer's Disease. *Alzheimer's Dement.* **2018**, *14*, 535–562. [CrossRef]

47. Jadon, S. SSM-Net for Plants Disease Identification in Low Data Regime. In Proceedings of the 2020 IEEE/ITU International Conference on Artificial Intelligence for Good, AI4G 2020, Virtual Event, 21–23 September 2020.

48. VGG16—Convolutional Network for Classification and Detection. Available online: https://neurohive.io/en/popular-networks/vgg16/ (accessed on 28 July 2022).
49. Classification: True vs. False and Positive vs. Negative | Machine Learning | Google Developers. Available online: https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative?hl=en (accessed on 28 July 2022).
50. GitHub. Tensorflow/Models: Models and Examples Built with TensorFlow. Available online: https://github.com/tensorflow/models (accessed on 17 August 2022).
51. GitHub. Heartexlabs/LabelImg: LabelImg Is a Graphical Image Annotation Tool and Label Object Bounding Boxes in Images. Available online: https://github.com/heartexlabs/labelImg (accessed on 28 July 2022).
52. Premium Photo. First-Person View of a Guy Eating Food in a Summer Cafe. Available online: https://www.freepik.com/premium-photo/first-person-view-guy-eating-food-summer-cafe_7862921.htm (accessed on 28 July 2022).
53. Peng, J.; Kang, S.; Ning, Z.; Deng, H.; Shen, J.; Xu, Y.; Zhang, J.; Zhao, W.; Li, X.; Gong, W.; et al. Residual Convolutional Neural Network for Predicting Response of Transarterial Chemoembolization in Hepatocellular Carcinoma from CT Imaging. *Eur. Radiol.* **2020**, *30*, 413–424. [CrossRef]
54. Kaggle. Fruit Images for Object Detection. Available online: https://www.kaggle.com/datasets/mbkinaci/fruit-images-for-object-detection (accessed on 28 July 2022).
55. Gill, K.; Mangat, V. Comparative Analysis of Deep Learning Methods for Object Detection. *Adv. Math. Sci. J.* **2020**, *9*, 3759–3775. [CrossRef]
56. Liu, Y.; Ma, Z.; Liu, X.; Ma, S.; Ren, K. Privacy-Preserving Object Detection for Medical Images with Faster R-CNN. *IEEE Trans. Inf. Forensics Secur.* **2019**, *17*, 69–84. [CrossRef]
57. Computer Vision—Dive into Deep Learning 1.0.0-Alpha0 Documentation. Available online: https://d2l.ai/chapter_computer-vision/index.html (accessed on 17 August 2022).
58. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
59. Activity Recognition—Wikipedia. Available online: https://en.wikipedia.org/wiki/Activity_recognition (accessed on 28 July 2022).
60. Kay, W.; Carreira, J.; Simonyan, K.; Zhang, B.; Hillier, C.; Vijayanarasimhan, S.; Viola, F.; Green, T.; Back, T.; Natsev, P.; et al. The Kinetics Human Action Video Dataset. *arXiv* **2017**, arXiv:1705.06950.
61. del Rose, M.S.; Wagner, C.C. Survey on Classifying Human Actions through Visual Sensors. *Artif. Intell. Rev.* **2012**, *37*, 301–311. [CrossRef]
62. Anand, P.; Singh, A.K.; Srivastava, S.; Lall, B. Few Shot Speaker Recognition Using Deep Neural Networks. *arXiv* **2019**, arXiv:1904.08775.
63. Hara, K.; Kataoka, H.; Satoh, Y. Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet? In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.
64. Rosebrock, A. *Human Activity Recognition with OpenCV and Deep Learning*; PyImageSearch: Philadelphia, PA, USA, 2019.
65. Carreira, J.; Zisserman, A. Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 4724–4733. [CrossRef]