Classifying Cafeteria Meals

Nicas Frank, Elias Teoman Eroglu, Joshua Nerling, Hannes Simon, Benedikt Schuster

Leipzig, 03.02.2025

Table of Contents

01

Business Case & Methodology

Business Case

Idea

New Payment Option

Self-checkout tills and smartphone app

Simple and Intuitive Use

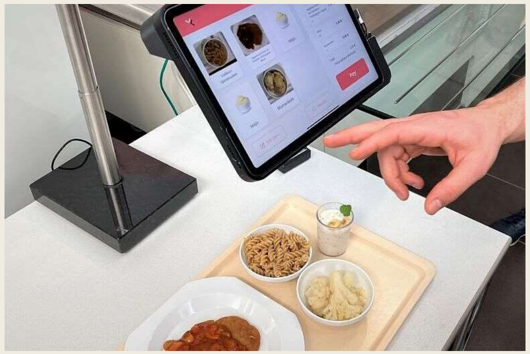
Automatic payment by scanning the dish

Basis: Machine Learning

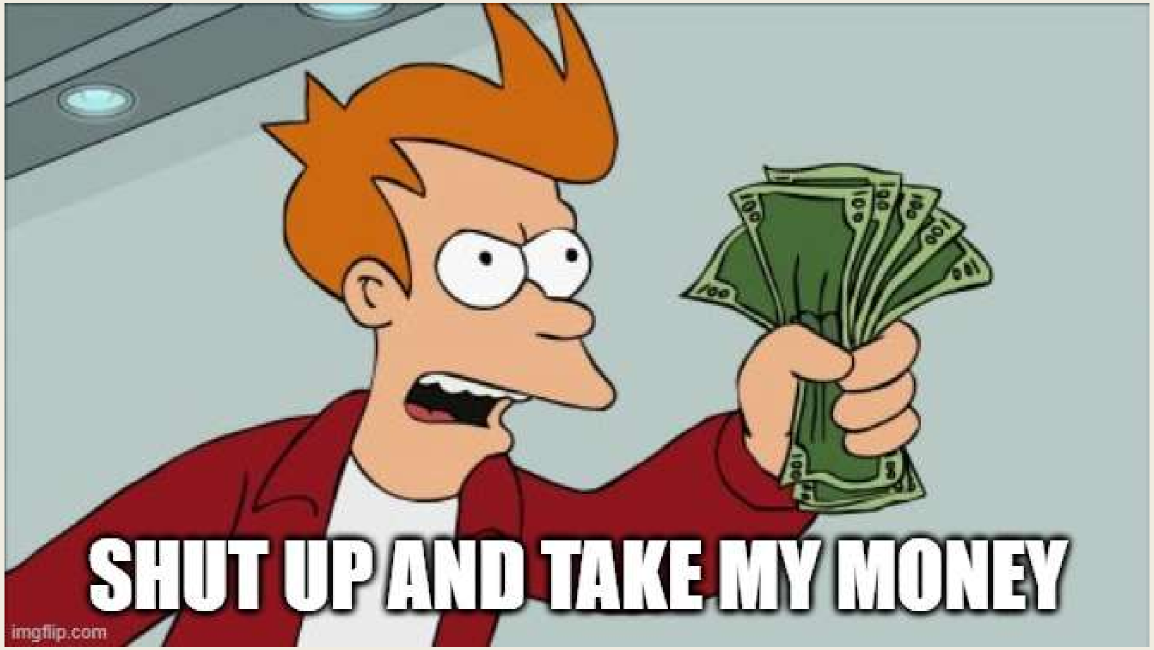
Models reliably classify the dishes

Benefits

1. Cost reduction
2. More efficient payment process
3. Customer satisfaction

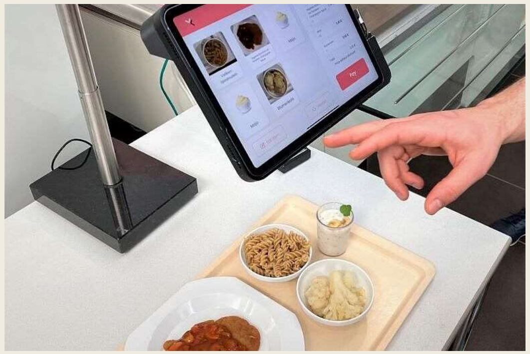


Business Case



Benefits

1. Cost reduction
2. More efficient payment process
3. Customer satisfaction



Approach to Solutions

CNN developed and trained by us

- Lean architecture
- Full control and transparency

- Limited data basis
- Higher risk of overfitting

EfficientNet with transfer learning

- Proven network
- Comprehensive data basis

- Higher complexity
- Limited transparency

Methodology

01

Definition of goals and success factors

02

Data collection

03

Data processing (preprocessing, data augmentation)

04

Model selection and initialization

05

Training process & model optimization

02

Evaluation of the Results & XAI

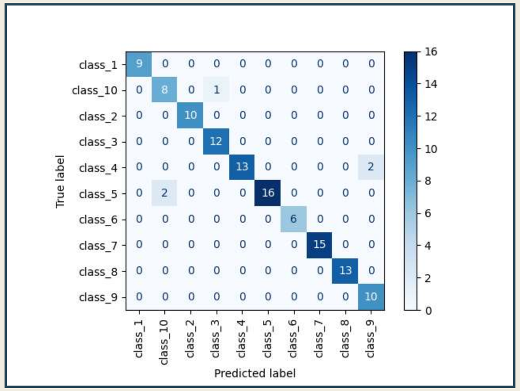
Comparing Model Results

	Custom CNN	EfficientNet (ImageNet)	EfficientNet (Food101)
Accuracy	0.92	0.96	0.99
Precision	0.92	0.95	0.99
Recall	0.93	0.96	0.99
F1-Score	0.92	0.95	0.99
Inference time	0,087 s	0,231 s	0,238 s
Parameters	25,71 million	6,53 million	6,53 million
GFLOPs	9,61	0,41	0,41

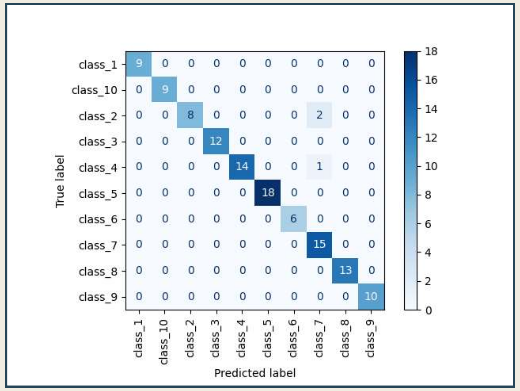
Comparing Model Results

	Custom CNN	EfficientNet (ImageNet)	EfficientNet (Food101)
Accuracy	0.92	0.96	0.99
Precision	0.92	0.95	0.99
Recall	0.93	0.96	0.99
F1-Score	0.92	0.95	0.99
Inference time	0,087 s	0,231 s	0,238 s
Parameters	25,71 million	6,53 million	6,53 million
GFLOPs	9,61	0,41	0,41

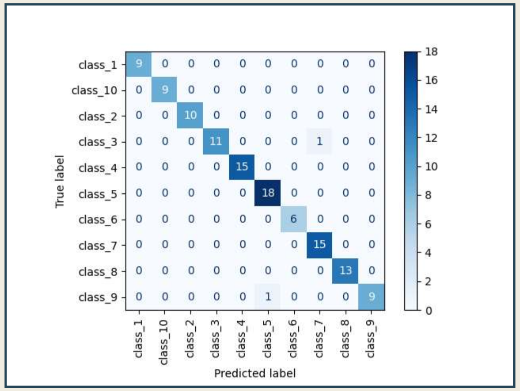
Comparing Model Results



Custom CNN



EfficientNet (ImageNet)

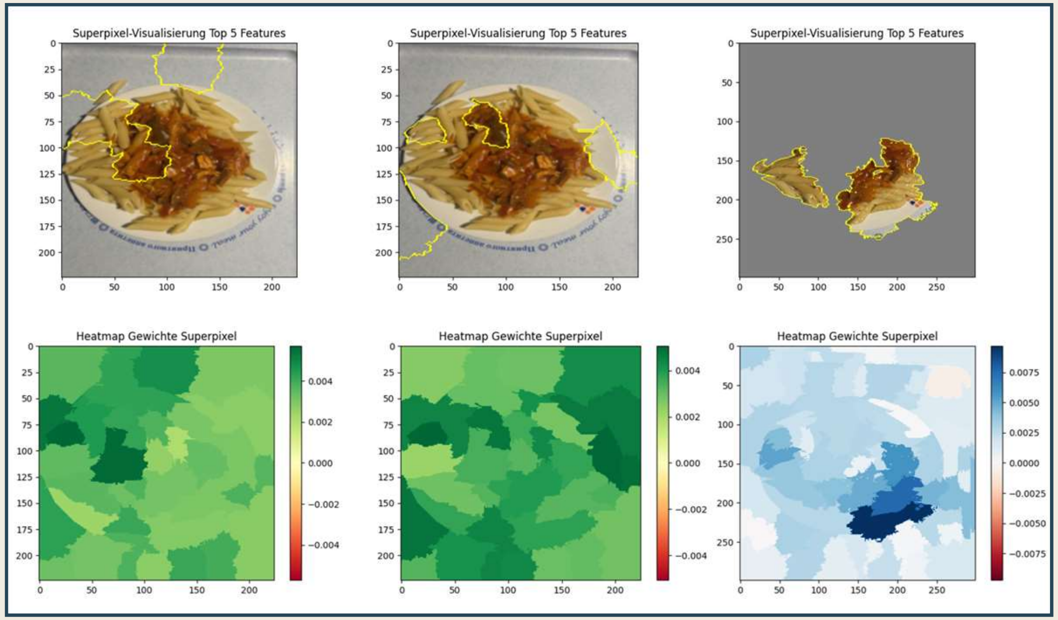


EfficientNet (Food101)

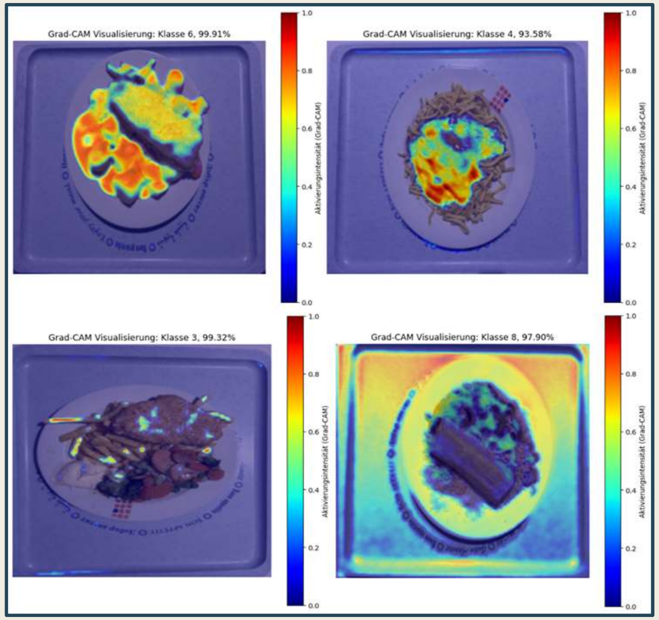
1. No prominent patterns of systematic errors are observed.
2. Predictions are well-distributed across classes, indicating strong generalization across dish categories.
3. No single class showing a significant number of misclassifications or consistent error trends.

XAI

LIME



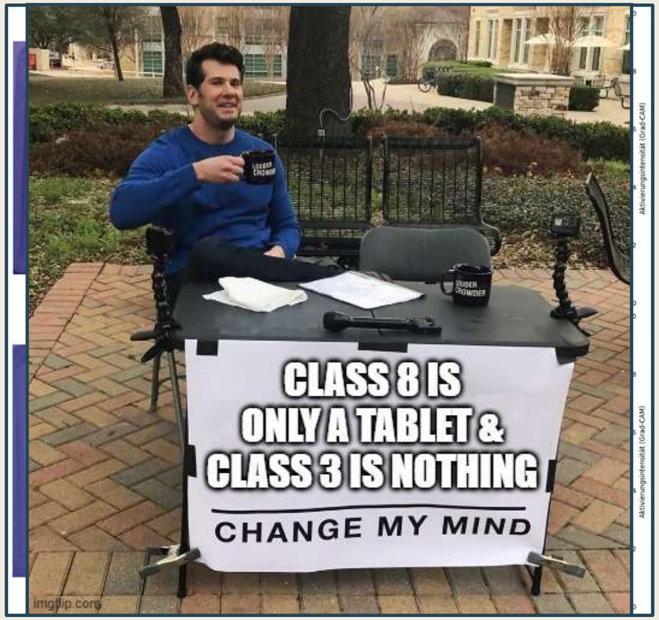
Grad-CAM



XAI

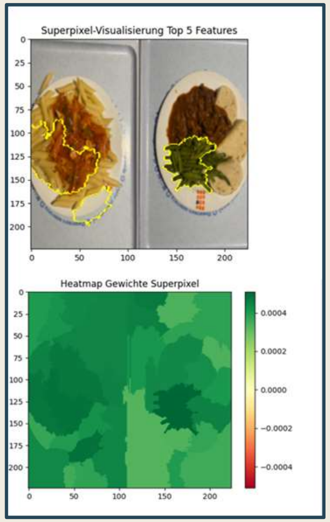
LIME

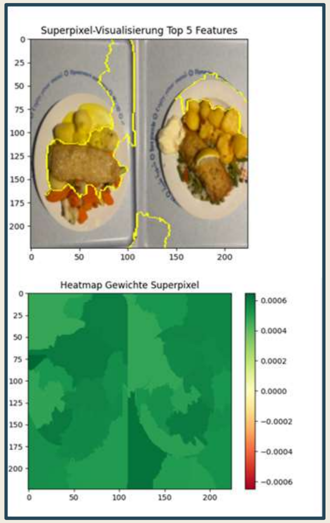
Grad-CAM

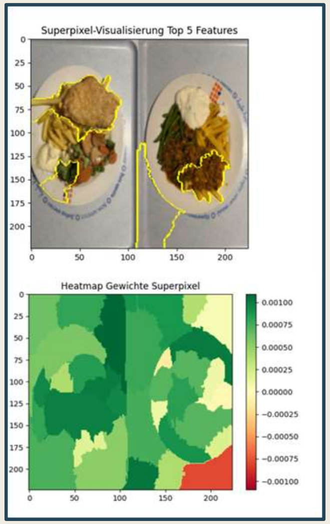


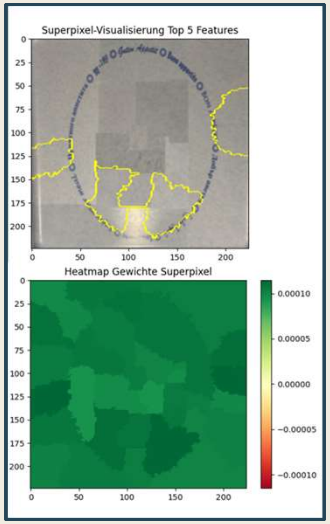
XAI

LIME

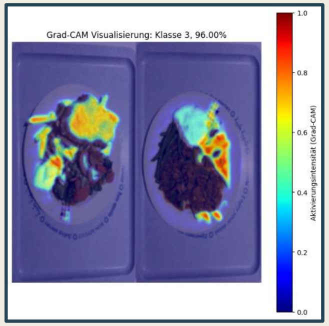


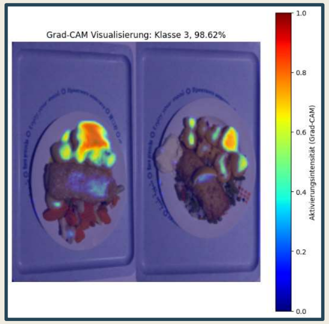


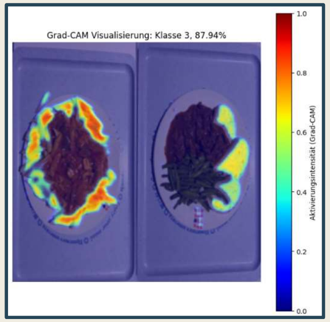


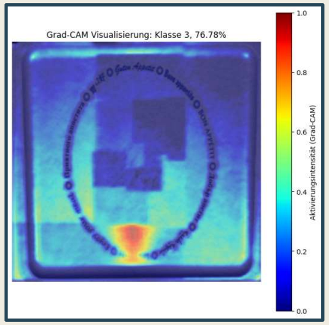


Grad-CAM





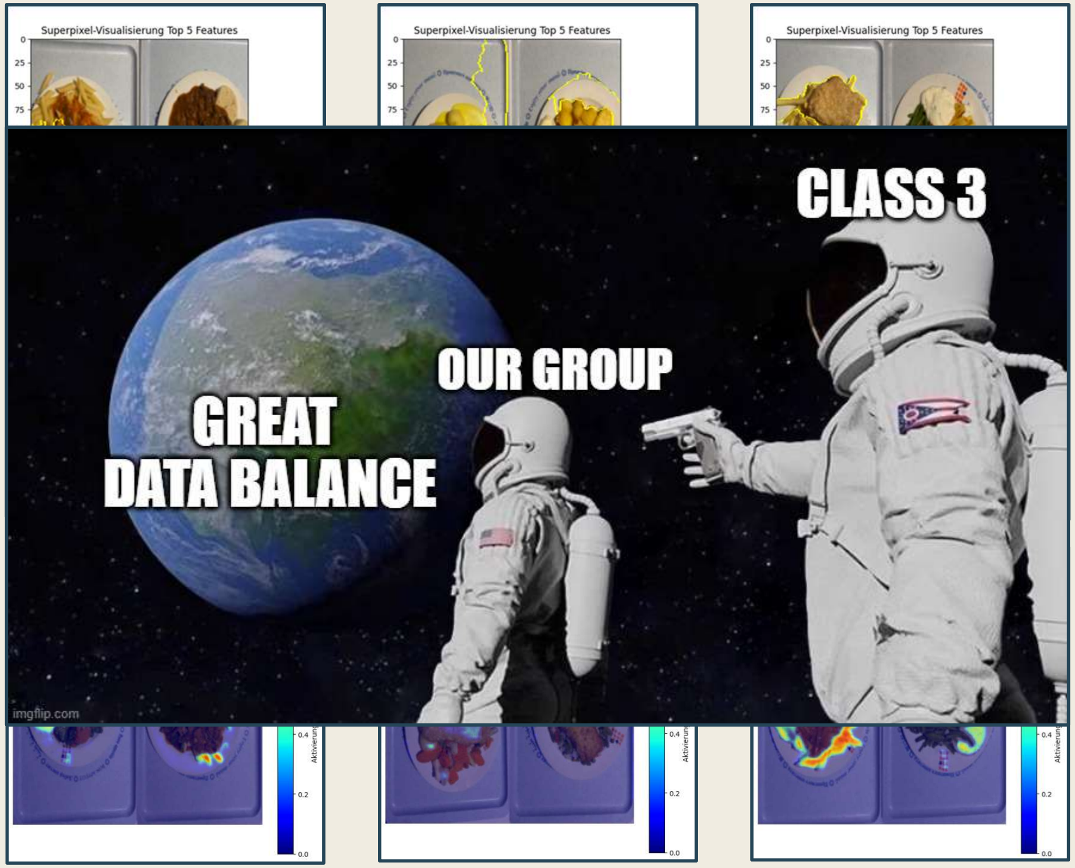


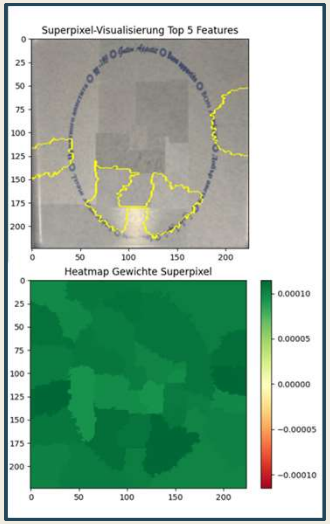


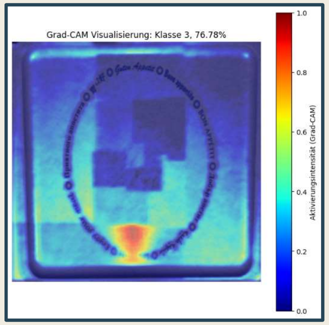
XAI

LIME

Grad-CAM







03

Critical Review & Outlook

Critical Review

Result Quality

- Promising Accuracies for all examined models
- Contractionary XAI-Results
- Risk of overfitting due to a controlled training environment

Data Collection

- Challenges with meal presentation variability
- Varying camera angles & lighting conditions
- Small Train/Valid-Dataset

Model Design

- Current classification works reliably only when a single dish is on the tray.
- High training effort required when introducing new dishes.

Business Case

- Mobile App vs. Fixed kiosks
- Fraud prevention

Future Work

Result Quality

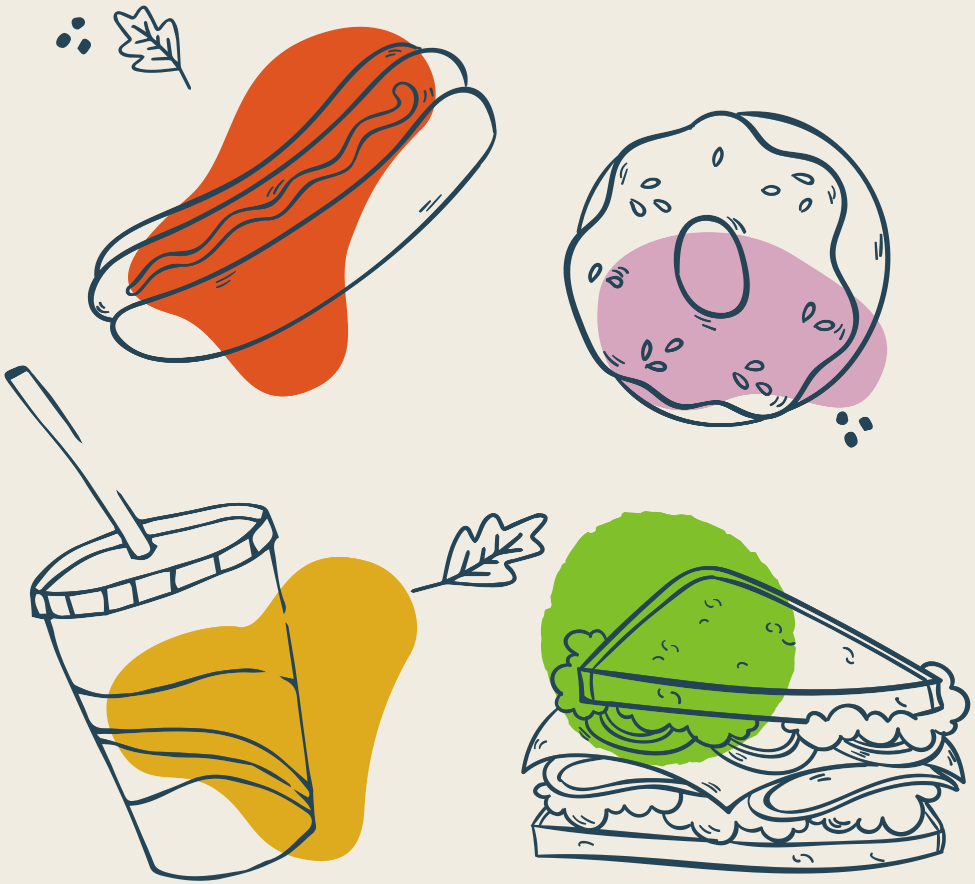
Data Collection

Model Design

Business Case

Suggestions

1. Expand Dataset Diversity
2. Integrate Object Detection
3. Conduct Comprehensive Real-World Testing
4. Explore Single-Model Approach for Each Meal
5. Enhance Mobile App with QR-Code Verification

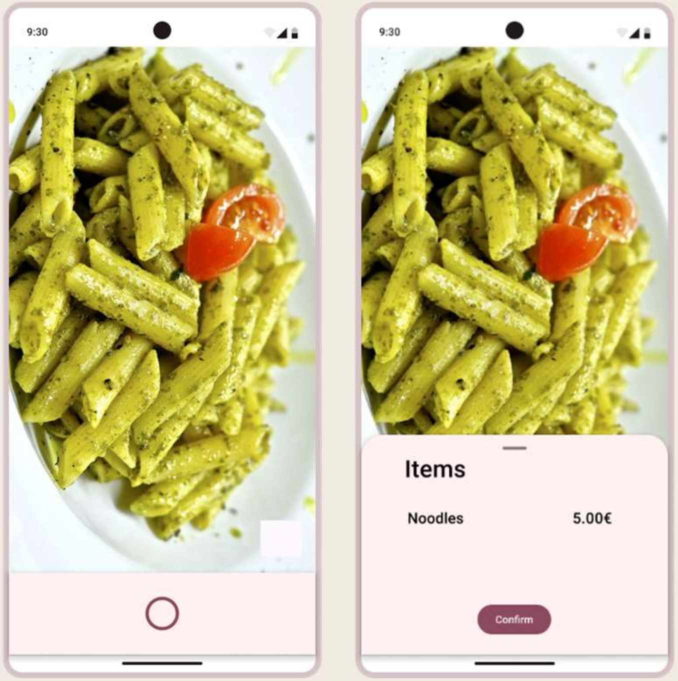


04

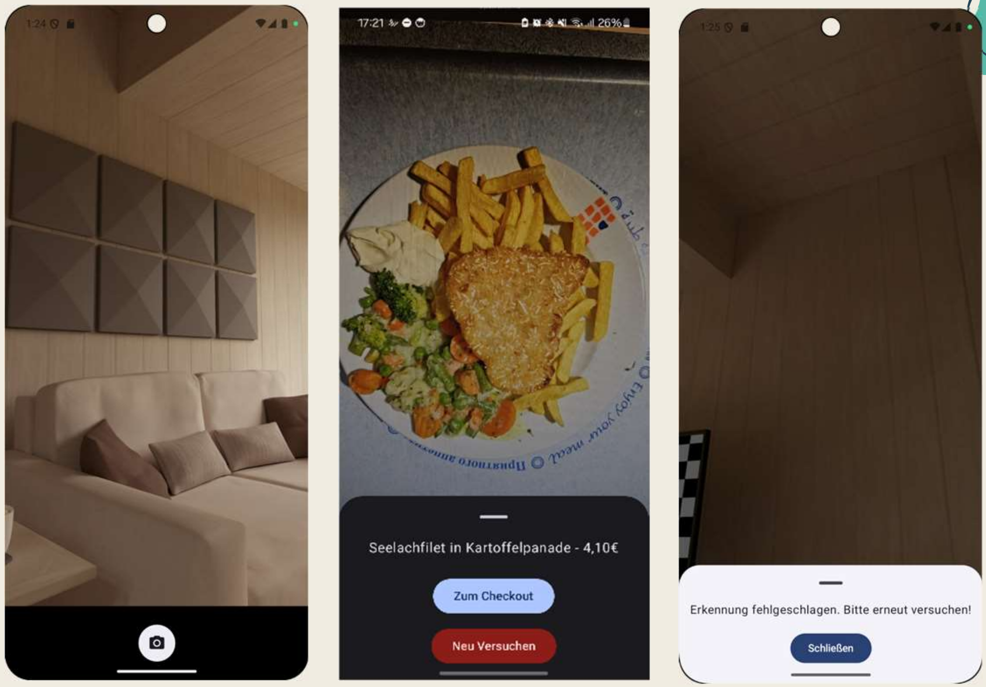
Model Deployment & Live Demo

Model Deployment

First Mockup



UI of MensAI



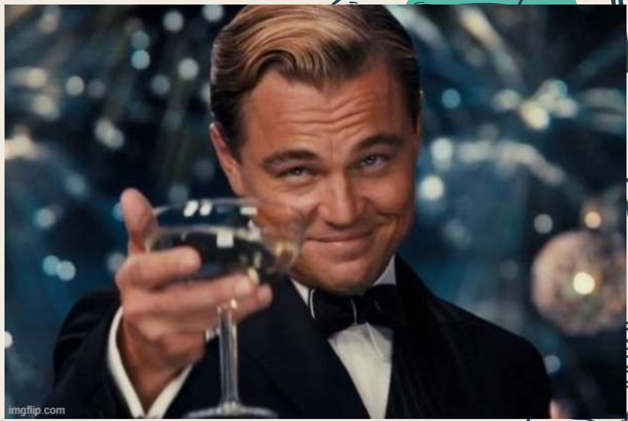
Live Demo of MensAI

Thanks!

Do you have any questions?

Nicas Frank, Elias Teoman Eroglu, Joshua Nerling, Hannes Simon, Benedikt Schuster

dr65cehe@studserv.uni-leipzig.de



References

- https://www.studierendenwerk-goettingen.de/fileadmin/_processed_/c/e/csm_NEWS-Selbstscan-Kasse_1__371c57ff26.jpg
- https://tutor-church-15580.netlify.app/assets/img/lime_logo.jpg
- https://miro.medium.com/v2/resize:fit:1162/0*pUaDvGZV8Q-5ZBhC.jpg
- Memes erstellt mit: https://imgflip.com/

05

Backup

Methodology

01 Definition of goals and success factors

02 Data collection

03 Data processing (preprocessing, data augmentation)

04 Model selection and initialization

05 Training process & model optimization

Definition of goals and success factors

Goals and success factors

- Precise detection of the dishes
- Ensure robustness under realistic conditions
- Transparency of model decisions through XAI
- Minimization of misclassification
- Scalability and efficiency
- Simple and self-explanatory operation

Metrics

- Accuracy
- Precision
- Recall
- F1-Score
- Inference time
- FLOPs
- Confusion matrix
- XAI visualizations

Data collection

Aim of the data procurement

- Ensuring a sufficient amount of data for classification
- Covering different variations of the dishes (e.g. different angles, lighting conditions, arrangement)

- 10 different dishes
- 50 original images per dish
- Slight variation of the arrangement in each picture

➢ Variations allow us to achieve a high level of data diversity and train the models for realistic scenarios

Data processing (preprocessing, data augmentation)

1 Division of the dishes into 10 classes

2 Data split within each class

- 64 % training data
- 16 % validation data
- 20 % test data

3 Data transformation & data extension

- Training data:
  - RandomResizedCrop, RandomHorizontalFlip, GaussianBlur, ToTensor, Normalize
- Test and validation data:
  - Resize, CentercropToTensor, Normalize

Class 1

Class 2

Class 3

…

…

…

Class 10

Data processing (preprocessing, data augmentation)

1 Random-Resized Crop

- Randomly cuts out an area and scales it to 224x224 pixels
- Goal: Increases variety and robustness against different plate positions

2 Horizontal Flip

- Mirrors the image horizontally at random
- Objective: Provides robustness against symmetrical variations (e.g. left-right rotations)

3 Gaussian blur

- Applies random blurring, simulates blurred images
- Objective: Trains the model to be robust against blurring

- **Increase the number of training images per class**: from 40 to 40 X number of epochs
- **More robust training**: Better generalization to new data
- **Better model performance**: Higher accuracy through more versatile training data

Architecture of the Convolutional Neural Network

01 02 03 04 05

1. INPUT LAYER
   - Batch-Size: 16
   - Input: (batch_size,3,224,224)

2. CONVOLUTION LAYER 1
   - Filter: 32
   - Kernels: 3x3
   - Padding: 1
   - Output: (batch_size, 32, 224, 224)

3. MAX-POOLING LAYER 1
   - Kernels: 2x2
   - Stride: 2
   - Output: (batch_size, 32, 112, 112)

4. CONVOLUTION LAYER 2
   - Filter: 64
   - Kernels: 3x3
   - Padding: 1
   - Output: (batch_size, 64, 112, 112)

5. MAX-POOLING LAYER 2
   - Kernels: 2x2
   - Stride: 2
   - Output: (batch_size, 64, 56, 56)

6. FLATTENING
   - Flattening the feature map
   - Output: (batch_size, 64 * 56 * 56)

7. FULLY CONNECTED LAYER 1
   - Neuronen: 128
   - Output: (batch_size, 128)

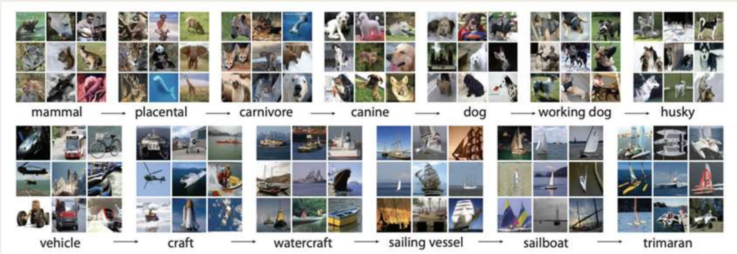
8. DROPOUT
   - 20 % probability of removal

9. FULLY CONNECTED LAYER 2
   - Neurons: 10 (Number of classes)
   - Output: (batch_size, 10) (Logits)

10. OUTPUT LAYER
   - Prediction = Maximum logit from pre-shift

Transfer Learning: EfficientNet

- Transfer learning enables the reuse of pre-trained neural networks for specific tasks.

	EfficientNet (ImageNet)	EfficientNet (Food101)
Pretrained on	ImageNet data set (general purpose images)	Food101 dataset (food)
		
Usage	Broad visual feature selection for general tasks	Specific feature selection for food classes
Modification	• Classifier layer for target class replaced • All previous layers frozen (no update of weights)	

Quellen:
- https://trendskout.com/wp-content/uploads/2020/04/image_recognition_illustration.png
- https://production-media.paperswithcode.com/datasets/Food-101-0000000037-8c457091_ZXHhL3x.jpg

01 02 03 04 05

Training process & model optimization

Optimization methods

1 Hyperparameter:

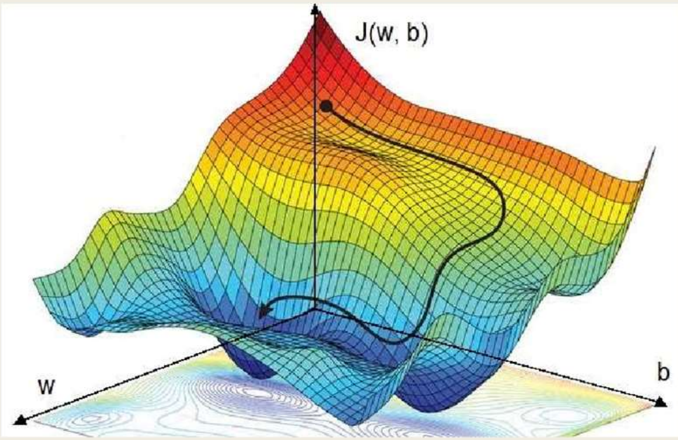
- Batch size: 16
- Learning rate: Determined by Learning Rate Finder
- Epochs: Maximum of 50 training epochs

2 Manual fine adjustment:

- Batchsize
- Model architecture
- Dropout

3 Adaptive learning rate

Early Stopping

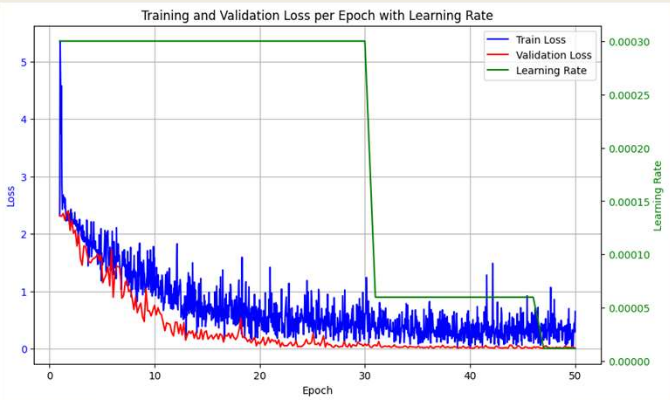


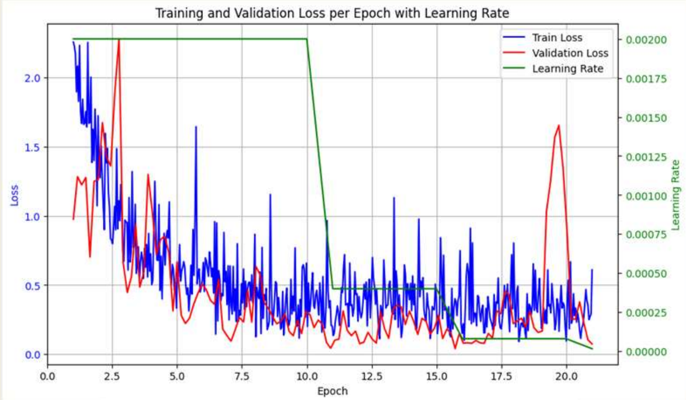
Quelle: Patlatzoglou, Konstantinos. (2022). Deep Learning for Electrophysiological Investigation and Estimation of Anesthetic-Induced Unconsciousness

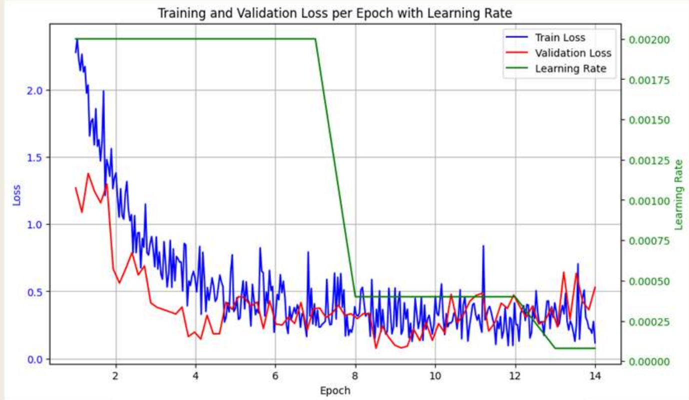
01 02 03 04 05

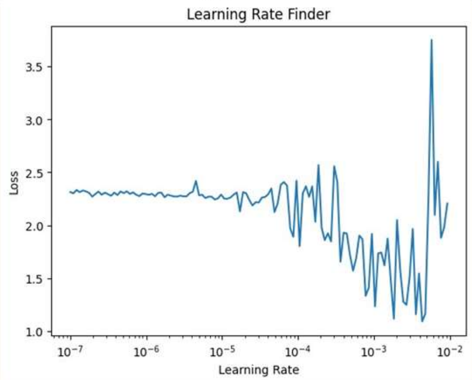
Training process & model optimization

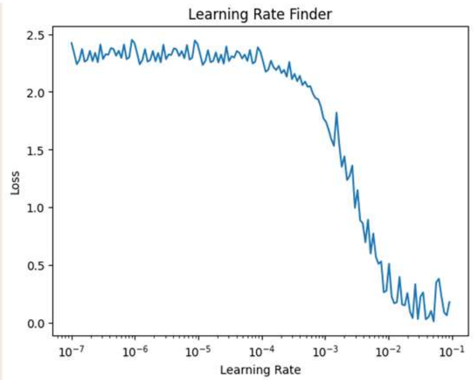
Convolutional Neuronal Network	EfficientNet (ImageNet)	EfficientNet (Food101)

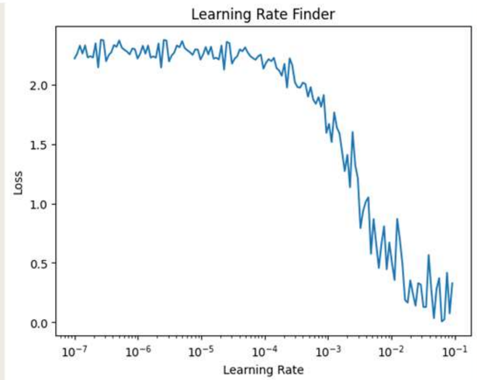




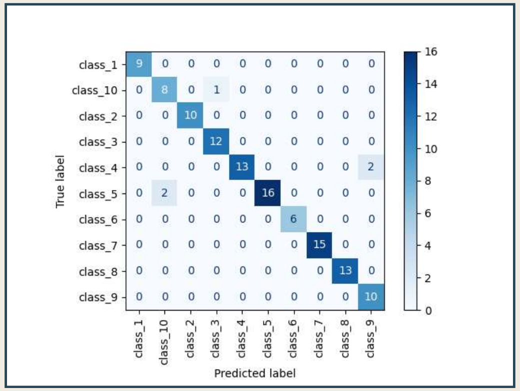




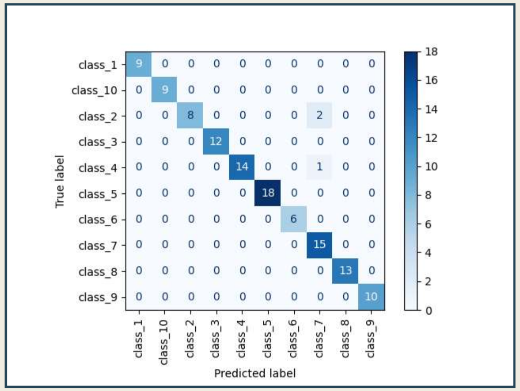




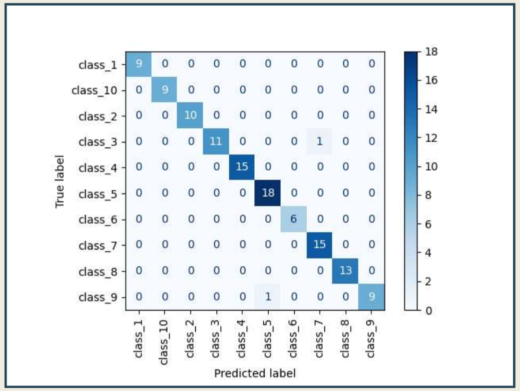
Comparing Model Results



Custom CNN



EfficientNet (ImageNet)



EfficientNet (Food101)

1. No prominent patterns of systematic errors are observed.
2. Predictions are well-distributed across classes, indicating strong generalization across dish categories.
3. No single class showing a significant number of misclassifications or consistent error trends.