Deep Learning-Based Mood Recognition in Dementia Patients: A Comparative Study of 2D and Thermal Imaging

Somaiya Abdulrahman* Sadeen Alkhalili† Ebba Norlin‡ Anas Mustafa Mohadin§ Valento Bardhoshi§ ¶
School of Innovation, Design and Engineering, M.Sc.Eng Robotics
Mälardalens University, Västerås, Sweden
Email: *san21025@student.mdu.se, †sai25005@student.mdu.se, ‡enn21010@student.mdu.se,
§amn21018@student.mdu.se, ¶vbi24001@student.mdu.se

***Abstract*—Currently, there is a big shift towards establishing treatment of behavior and psychological symptoms in dementia in non-pharmacological ways. This paper investigates the use of 2D and thermal images for mood detection using deep learning models. Experiments were conducted using the two publicly available data sets, KFTE and TUFT. The datasets contain both thermal and 2D images and vary in ethnicity, age, gender, etc. Two models containing a convolutional neural network, ResNet50 and EfficientNet-B0, were used to analyse and classify the emotional states of the images. Each model was trained and tested on each dataset to give results of thermal image and 2D image separately, late fusion was used to compare the results with only using 2D images and thermal images. The results showed that using late fusion improved the overall accuracy, showing that combining 2D image together with thermal image may be used as a non-pharmacological way to treat behavior and psychological symptoms in dementia.**

***Index Terms*—Mood detection, Convolution Neural Network - CNN**

I. Introduction

Over the years the ability to analyze and detect human emotions using artificial intelligence has been extensively researched over the years. The applications of detecting emotions range from improving security to healthcare. An important area is dementia care, where experiential recognition can be crucial in understanding and managing symptoms associated with the disease.

As of 2021, 57 million people around the world had been diagnosed with dementia, with approximately 10 million new cases each year [1]. Dementia is a term used to classify a group of diseases that over time destroy nerve cells in the brain, resulting in impaired cognitive function[2]. The cognitive function of the brain is the ability to process our thoughts, managing mood, emotional control, memory, behaviour and motivation. Majority of the patients with dementia experience Behavioural and Psychological Symptoms of Dementia, BPSD episodes[3]. The patient often feels hopelessness, guilt, fear, anxiety or anger. The result contributes to dementia patients having outbursts when not understanding their own emotion or situation [4]. Furthermore, due to the impaired cognitive function, patients with dementia may struggle to express their emotions or understand their mental state[5]. Therefore, the caregiver needs experience and awareness to be able to handle BPSD episodes. Since these episodes occur unpredictably they can be harmful on a both physical and psychological level to both the patient and caregiver[5]. The outcomes of a patient experiencing BPSD episodes without proper care are shame, longer hospital stays, signed into caring homes at earlier stages, suicidal thoughts and declined mental health[4]. One way to improve the provided care for dementia patients would be to prevent BPSD episodes before they happen[5]. By using mood detection technology to capture mood changes to better predict BPSD episodes, caregivers have a higher chance of providing the right care at the right time[6]. 2D images and CNN have previously been used to detect human emotion. However, the possibility of using multi-modal with late fusion could lead to higher precision[7].Thermal imaging, scanning the patients skin temperature, could further provide data to improve precision in emotion prediction. Especially, since patients with dementia do not always express emotions in predictable ways[5]. This paper aims to further investigate the possibility of detecting mood changes and emotions in dementia patients using 2D and thermal imaging, making it possible to better understand and predict BPSD episodes.

II. Background

A. Image processing using CNN

Image processing using an CNN, convolutional neural network, is one of the most well used architectures for

image processing[8].The CNN is inspired to work similar to the way human vision and mind processes images. The images are processed through three different layers that helps the model extract hierarchical features[9].

CNN is a neural network that is feedforward and has the ability to extract information from data with convolution structures [10]. The architecture of CNN use three key features[10].

1) **Local connections:** There is no connection between a neuron and all the neurons of preceding layer, instead each neuron is connected to some neurons. This results in reduced parameters and improved convergence.
2) **Weight sharing:** Weight sharing means a group of connected neurons have the ability to share the same weights. Reusing the same filters (kernels) over the entire image which leads to further reduced parameters.
3) **Downsampling (Pooling):** Downsampling through pooling makes it possible to retain important information from images while reducing the size of the image. This also reduces the number of parameters, as less important information is disregarded.

Techniques such as convolutional neural networks (CNNs) have shown promising results in classifying different emotions such as happiness, anger, fear, confusion, surprised, and sadness [11], [12], [13], [14]. Two common used CNN-based architectures used for emotion detection are EfficientNet and ResNet.

1) EfficientNet-B0: EfficientNet, a model constructed and introduced by Mingxing Tan and Quoc V.Le in 2020 [15]. EfficientNet use the compound scaling method to balance the three dimensions of CNN. Resolution (input image size), depth (number of layers) and width (number of channels) is balanced using fixed scaling coefficients. A image of higher resolution needs a deeper network to properly capture details in the image[16]. The EfficientNet therefore achieves high accuracy while remaining computationally efficient, making it suitable for real-time applications.

2) ResNet-50: ResNet is a CNN model that use residual learning that solves the problem with the vanishing gradient in deep networks. The vanishing gradient becomes a problem when the gradients become very small during backpropagation, causing training mistakes[17]. The ResNet model solves this problem with skipped connections, which allows the network to bypass certain layers and directly pass information forward[18].

B. 2D-Imaging

One of the most widely explored methods for detecting emotion in facial expression analysis is using 2D images. Two-dimensional or briefly known as 2D is an object that is virtual with no depth [19]. 2D imaging can be used for mood detection by extracting the features of the face and train a CNN to classify the different emotions.

C. Thermal Imaging

Thermal imaging, according to [20], is a technique that uses infrared technology to recognize heat emission from different objects. This unique process converts the infrared energy (IR) to visible display. Thermal images provides a heat map of an area, giving more information than a normal 2D image. Using thermal imaging in emotion detection is relevant as the different parts of the face will have different temperatures based on the emotional state. Embarrassment increases blood flow to cheeks, blushing, making the skin red and hot [21]. While stress can be portrayed by sweating and heating around the mouth.[22]

III. Ethical Consideration

The application of AI in health- and patient care brings important ethical challenges into light[23]. In this specific case, which involves patients with dementia, the patients have a cognition impairment. Different aspects must be taken into consideration, such as ensuring that patients have the ability to give informed consent to participating in a study and make fully informed decisions about being monitored by providing PII data[24]. The patients have to be comfortable with, as well as give permission, for caregivers to have access to and process their personal and health-related information[24].

By monitoring patients' facial expressions using thermal- and 2D images to detect their emotional and physiological states. To give proper care adjusted to each individual patient, data may need to be collected continuously, which requires active participation. This raises the question whether it is ethically acceptable to continuously monitor vulnerable individuals[24].

Lastly, the method for collecting data and who was the authority over it comes with ethical challenges. Challenges include to determine who controls the use of the data and who has the authority to make decisions based on the mood detection outputs. In this case, the patient, caregivers, or medical staff[24].

IV. Related Work

Over the years, advances in machine learning, especially CNNs, have made it possible to accomplish things such as facial recognition from images [10]. There are several methods of detecting emotions from facial images. In this section, studies using different methods of emotion recognition will be presented.The insights from these studies will be used to aid this paper in finding answers to the proposed research questions.

A. Thermal imaging

Several experiments have been performed on emotion detection using facial skin temperature. Kahil Mustafa Jamal and Eiji Kamioka [25] showed promising results when it comes to the detection of different human emotions using facial skin temperature together with heart rate variability. Although there exist several unique emotions, the study focused on four emotions (relax, fear, joy, and sadness). With the cross-validation method (ANN), the accuracy of the estimate achieved was 88. 75%.

Ilikci et al. [26] conducted a study in which they used different fast detection algorithms and compared them to classify different emotions. Thermal imaging was used in [26], the reason being that there are ways to hide emotions and provide false information. With the help of heat maps, given by thermal imaging, the problem can be weakened. The three different algorithms chosen were YOLOv3, ResNet, and DenseNet.The research showed the possibility of using thermal images in emotion detection.

B. 2D-imaging

Bilal Taha and Dimitrios Hatzinakos [27] did a research regarding emotion detection from 2D images. They developed their own CNN model and tested it against different extraction techniques, Local Binary Pattern (LBP), Laplacian of Gaussian (LoG) and a fused version which is a combination of the two. The results show that the fused version got higher accuracy then LBP and LoG individually, most notably the CNN model got the highest accuracy.

Another study were they created their own CNN model was done by Sakshi Gupta and Anwesha Sengupta [28]. The CNN model was compared with two already existing models namely YOLOv5 and YOLOv5-NMS. The proposed method received average of 98.57% accuracy in detecting emotion from the given data set.The results show the possibility of emotion detection from thermal images using created CNN and existing models.

Khajontantichaikun et al. [29] did a research where they compared different object detecting algorithms regarding facial emotion detection. They trained and tested the models with a dataset containing Thai elderly people. The three different models used were YOLOv7, Faster R-CNN and SSD. The result showing the possibility of using CNN even with dataset containing elderly people.

Thomas et al. [30] conducted a study where they used a model with CNN to detect emotions from images. The model was used to recognize real-time emotions and categorize them into 7 different emotions.The results showed that the created model that contained CNN could with success detect the 7 different emotions (sad, fear, happy, angry, neutral, surprised and disgust) with the accuracy of 74%.

C. EfficientNet

A study were they utilized Efficientnet was done by Yagan Arun and Viknesh G S [31]. They did leaf classification for plant recognition using the Efficientnet architecture. The study showed the possibility of modifying EfficientNet to detect and classify new data.

V. Problem formulation

The following research questions where conducted:

1) ***What is the accuracy, precision, recall, and F1-score of classifying (emotional states/unsettling feelings) using a CNN-based model trained on healthy adult data containing 2D- and thermal images respectively?***
2) ***How does combining 2D and thermal images (sensor fusion) impact the accuracy of emotional state classification compared to using each model individually?***

VI. Method

A. Data collection

Although the aim of the project is to test emotion detection tools on dementia patients, it has been challenging to obtain and collect a dataset that includes emotions from this specific group. As a result, facial images has been collected from already existing databases that includes people of different ages.

In this study, two different datasets were used, KTFE[32] and TUFT[33].

Subjects (participants)	26 (10 females + 16 males)
Classes (emotions)	7
Countries	Vietnam, Japan, Thailand
Age Range	11–32 years
Total Images	2,252

Table I: Participant demographics for the KTFE dataset

Emotion	Training	Validation	Total
Anger	286	65	351
Disgust	106	48	154
Fear	129	110	239
Happy	493	128	621
Neutral	62	17	79
Sad	488	123	611
Surprise	158	39	197
Total	1,722	530	2,252

Table II: Class distribution in the KTFE dataset (training and validation split)

Subjects (participants)	112 (74 females + 38 males)
Classes (emotions)	4
Countries	15+
Age Range	4–70 years
Total Images	444

Table III: Participant demographics for the TUFT dataset

Emotion	Training	Validation	Total
Neutral	89	22	111
Happy	89	22	111
Sleepy	89	22	111
Surprise	89	22	111
Total	356	88	444

Table IV: Class distribution in the TUFT dataset (training and validation split)

B. Model tuning

The EfficientNet-B0 and ResNet-50 models was implemented in python language using the Pytorch library. The models are publicly available and pre-trained with the imageNet dataset. The models for this project are adjusted to classify emotions instead of objects.

C. Implementation

The KTFE and TUFT data sets were used as the source of facial images for 2D and thermal images.

1) Data augmentation and preprocessing: To make sure the diversity of the dataset, several preprocessing steps where conducted to prevent potential overfitting. Each image was cropped to the correct input size of 224x224 with the existing face recognition-tool MTCNN in python. Further augmentation of the training set was made by randomize the angle, contrast and lightning of the images.

2) Training process: Each model trained the 2D and thermal images separately with maximum of 50 epochs per training.The training cycle stops when the model shows signs of overfitting e.g. when there is no more improvement of the validation and loss variables. The

Model	Learning rate	Batch Size	Weight Decay
ResNet-50	0.001	64	0.0001
EfficientNet-B0	0.0001	20	0.00001

Table V: Training parameters for ResNet-50 and EfficientNet-B0

models are optimized with weight decay and learning rate to distribute the weights equally over the model. Due to the uneven classes in the KTFE dataset the training is adjusted to make sure each class is equally trained. To further reduce the risk of overfitting in the smaller classes.

3) Late fusion: The trained weights for 2D and thermal images are used to combine in late fusion. The late fusion combines the 2D and thermal images from the same person to output a combined prediction. Each prediction is the probability for each image in each class. The late fusion prediction is given by the following equation:

$$\text{Late Fusion Prediction} = \frac{2Dpred + Thermalpred}{2} \tag{1}$$

D. Evaluation

The performance of the models, ResNet50 and EfficientNet B0, is measured by the evaluation metrics (2)–(5) where True Positive (**TP**), False Positive (**FP**), True Negative (**TN**), False Negative (**FN**) represents the following in our study:

- TP refers to the cases where the model correctly predicts a specific mood class (e.g. predicting angry when the true label is angry).
- FP refers to the cases where the model incorrectly predicts a specific mood class when the actual mood is something else (e.g predicting angry when the true label is sad).
- TN refers to the cases where the model correctly identifies that a sample does not belong to a specific mood class (e.g not predicting angry when the true label is sad, fear, disgust).
- FN refers to the cases where the model fails to predict a specific mood class when it is the true label (e.g predicting sad when the true label is angry).

The accuracy measures the overall correctness of the model and is calculated as shown in Equation 2.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

Precision, as defined in Equation 3, measures the model's ability to correctly identify positive instances out of all instances it predicted as positive.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

Recall, as described in Equation 4, helps measure the model's ability to identify all relevant instances of a specific class. More specifically, it is the ratio of correctly predicted positive samples to all actual positive samples.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

Lastly, the F1-score, as shown in Equation 5, is caluclated. The F1-score is the harmonic mean of precision and recall and is a metric useful for cases when the class distribution is imbalanced.

$$F_1\text{-score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \tag{5}$$

VII. Result

A. *EfficientNet B0*

Emotion	Accuracy	Precision	Recall	F1-score
Neutral	0.95	0.95	0.86	0.90
Happy	0.95	0.85	1.00	0.92
Sleepy	1.00	1.00	1.00	1.00
Surprise	0.98	1.00	0.91	0.95

Table VI: EfficientNet B0 emotion classification performance of the model trained on 2D images from the TUFT dataset.

Emotion	Accuracy	Precision	Recall	F1-score
Neutral	0.88	0.74	0.77	0.76
Happy	0.88	0.74	0.77	0.76
Sleepy	0.93	0.81	0.95	0.88
Surprise	0.92	0.93	0.68	0.79

Table VII: EfficientNet B0 emotion classification performance of the model trained on Thermal images from the TUFT dataset.

Emotion	Accuracy	Precision	Recall	F1-score
Neutral	0.95	0.91	0.91	0.91
Happy	0.97	0.88	1.00	0.94
Sleepy	1.00	1.00	1.00	1.00
Surprise	0.97	1.00	0.86	0.93

Table VIII: EfficientNet B0 emotion classification performance of the model trained using late fusion of 2D and thermal images from the TUFT dataset.

Emotion	Accuracy	Precision	Recall	F1-score
Anger	0.98	0.52	0.65	0.58
Disgust	0.97	0.87	0.81	0.87
Fear	0.83	0.60	0.49	0.61
Happy	0.83	0.65	0.66	0.85
Neutral	0.97	0.57	0.70	0.57
Sad	0.75	0.46	0.39	0.46
Surprise	0.91	0.46	0.72	0.46

Table IX: EfficientNet B0 emotion classification performance of the model trained on 2D images from the KFTE dataset.

B. *ResNet50*

Emotion	Accuracy	Precision	Recall	F1-score
Neutral	0.94	0.87	0.91	0.89
Happy	0.97	0.88	1.00	0.94
Sleepy	0.97	1.00	0.86	0.93
Surprise	0.99	1.00	0.95	0.98

Table XII: ResNet-50 emotion classification performance on 2D images from the TUFT dataset.

Emotion	Accuracy	Precision	Recall	F1-score
Anger	0.92	0.74	0.57	0.64
Disgust	0.94	0.80	0.50	0.62
Fear	0.82	0.60	0.42	0.49
Happy	0.92	0.83	0.86	0.85
Neutral	0.98	1.00	0.29	0.45
Sad	0.76	0.49	0.68	0.57
Surprise	0.90	0.40	0.64	0.49

Table X: EfficientNet B0 emotion classification performance of the model trained on thermal images from the KFTE dataset.

Emotion	Accuracy	Precision	Recall	F1-score
Anger	0.92	0.67	0.68	0.67
Disgust	0.98	0.95	0.79	0.86
Fear	0.87	0.77	0.50	0.60
Happy	0.92	0.82	0.88	0.85
Neutral	0.99	1.00	0.65	0.79
Sad	0.80	0.55	0.73	0.63
Surprise	0.95	0.66	0.69	0.68

Table XI: EfficientNet B0 emotion classification performance of the model trained using late fusion of 2D and thermal images from the KFTE dataset.

Emotion	Accuracy	Precision	Recall	F1-score
Neutral	0.77	0.67	0.18	0.29
Happy	0.85	0.76	0.59	0.67
Sleepy	0.72	0.48	0.95	0.64
Surprise	0.89	0.76	0.73	0.74

Table XIII: ResNet-50 emotion classification performance on thermal images from the TUFT dataset.

Emotion	Accuracy	Precision	Recall	F1-score
Neutral	0.98	1.00	0.91	0.95
Happy	0.98	0.92	1.00	0.96
Sleepy	0.98	0.95	0.95	0.95
Surprise	0.98	0.95	0.95	0.95

Table XIV: ResNet-50 performance using late fusion of 2D and thermal images on the TUFT dataset.

Emotion	Accuracy	Precision	Recall	F1-score
Anger	0.46	0.05	0.20	0.08
Disgust	0.90	0.00	0.00	0.00
Fear	0.79	0.17	0.14	0.16
Happy	0.65	0.30	0.33	0.31
Neutral	0.96	0.13	0.06	0.08
Sad	0.61	0.20	0.20	0.20
Surprise	0.92	0.00	0.00	0.00

Table XV: ResNet-50 performance on 2D images from the KFTE dataset.

Emotion	Accuracy	Precision	Recall	F1-score
Anger	0.81	0.00	0.00	0.00
Disgust	0.85	0.06	0.04	0.05
Fear	0.65	0.12	0.11	0.11
Happy	0.77	0.53	0.33	0.41
Neutral	0.96	0.00	0.00	0.00
Sad	0.64	0.23	0.24	0.23
Surprise	0.70	0.09	0.36	0.15

Table XVI: ResNet-50 performance on thermal images from the KFTE dataset.

Emotion	Accuracy	Precision	Recall	F1-score
Anger	0.50	0.05	0.10	0.10
Disgust	0.90	0.06	0.06	0.06
Fear	0.76	0.15	0.13	0.12
Happy	0.72	0.41	0.38	0.39
Neutral	0.96	0.05	0.02	0.04
Sad	0.57	0.13	0.15	0.14
Surprise	0.89	0.12	0.08	0.10

Table XVII: ResNet-50 performance using late fusion (2D + thermal) on the KFTE dataset.

VIII. Discussion

A. Interpretation of results/Evaluation

The result show that the EfficientNet-B0 model outperform the ResNet-50 model on both datasets, especially the KTFE dataset. The result also show the potential of thermal images used in emotion detection. Since the thermal images could represent a more accurate emotion when the face may lie. As seen in X the emotions with lowest performance for KTFE, EfficientNet, are fear and sadness, potentially feelings that are difficult to feel in a safe environment. Further, it shows in the TUFT dataset when the result shows lower overall performance on thermal images than the 2D images. Since the TUFT subjects were instructed to show exaggerated emotion instead of induced to feel.

It also shows that the combined 2D and thermal images improves the overall performance of both models. Proving our research-question that a multimodal system improves the possibility of detecting accurate emotions.

B. Dataset impact

The KTFE dataset, which consists mainly of East and Southeast Asian individuals, displays subtle emotions and tends to suppress overt and exaggerated emotional expressions due to cultural norms. As a result, emotional changes in this dataset are often subtle and less exaggerated compared to datasets like TUFT, making it hard for the model such as ResNet-50 to detect and predict emotions. However, this type of dataset can be beneficial, especially considering that dementia patients commonly exhibit muted emotional expressions. Additionally, the KTFE dataset includes a wide range of labelled emotions, making it valuable for training and evaluating the models.

The TUFT dataset included fewer classes of emotions but more exaggerated emotions, resulting in a higher overall performance. Due to the wider range of age and ethnicity of the TUFT dataset it proves that both models is possible to use with a diverse population. Moreover, although TUFT includes fewer emotion classes, it shows the capacity for emotion detection when expressions are more visible.

C. Ethical Implications

The goal is to minimize potential harm to patients and improve care. Therefore the predicted emotion could impact the care given to the patients and have monumental consequences. In this case, high accuracy does not equal efficient prediction. It is also relevant to see that the ResNet-50 model is not performing high enough to be used in healthcare for now.

However, the EfficientNet-B0 model shows a higher performance over several emotions. The emotions often triggering BPSD episodes are unsettling feelings, such as angry, fear, sad, surprise and disgust. Besides disgust the EfficientNet-B0 has an average of 69% F1 score for those feelings. Not being precise enough to predict emotions in health care today as the results would be to unpredictable to rely on. However, shows potential for future work.

IX. Conclusion

In order to enable improved care for individuals with dementia, we examined in this research project how deep learning models can assist in detecting mood changes using 2D and thermal images. We tested two popular models on two datasets: ResNet50 and EfficientNet-B0. Our findings demonstrated that implementing late fusion to combine 2D and thermal pictures produced higher accuracy than using either one alone. In general, EfficientNet-B0 outperformed, particularly when it came to identifying subtle emotions. The ResNet50 model provided important insights into the variation in emotion detection performance, despite having trouble with several classes, particularly in the KTFE dataset. Further the result show the potential of thermal emotion detection due to it's inability to lie as the face could. This study supports the feasibility of multimodal mood detection as an alternative method to predict and maybe prevent BPSD episodes, despite several limitations, such as the dataset's generalizability to real dementia patients. Future research needs to tackle ethical issues such as data protection, continuous monitoring, and consent ,in addition to applying real-world validation on datasets specific to dementia.

REFERENCES

[1] World Health Organization, "Dementia," https://www.who.int/news-room/fact-sheets/detail/dementia, 2025, accessed April 16, 2025.

[2] A. Association, "What is dementia," https://www.alz.org/alzheimers-dementia/what-is-dementia, 2025, accessed June 8, 2025.

[3] V. Bränsvik, E. Granvik, L. Minthon, P. Nordström, and K. N. and, "Mortality in patients with behavioural and psychological symptoms of dementia: a registry-based study," *Aging & Mental Health*, vol. 25, no. 6, pp. 1101–1109, 2021, pMID: 32067466. [Online]. Available: https://doi.org/10.1080/13607863.2020.1727848

[4] N. Kar, "Behavioral and psychological symptoms of dementia and their management," *Indian Journal of Psychiatry*, vol. 51, no. Suppl 1, pp. S77–S86, Jan 2009.

[5] M. Healt, *Assesment and Management of Behaviour and Psychological Symptoms associated with Dementia (BPSD)-A Summary Handbook*, 1st ed., NSW Ministry of Health, NSW Ministry of Health 1 Reserve Road, December 2022, accesed: 2025-06-08.

[6] A. Siddiqui, P. Khanna, S. Kumar, and Pragya, "Progression analysis and facial emotion recognition in dementia patients using machine learning," in *Proceedings of International Conference on Network Security and Blockchain Technology*, J. K. Mandal, B. Jana, T.-C. Lu, and D. De, Eds. Singapore: Springer Nature Singapore, 2024, pp. 489–500.

[7] H. C. Qihang Yang, Yang Zhao, "Mmlf: Multi-modla multi-class late fusion for object detection with uncertainty estimation," 2024. [Online]. Available: https://arxiv.org/html/2410.08739v1

[8] F. Sultana, A. Sufian, and P. Dutta, "Advancements in image classification using convolutional neural network," in *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*. IEEE, Nov. 2018, p. 122–129. [Online]. Available: http://dx.doi.org/10.1109/ICRCICN.2018.8718718

[9] N. Sharma, V. Jain, and A. Mishra, "An analysis of convolutional neural networks for image classification," *Procedia Computer Science*, vol. 132, pp. 377–384, 2018, international Conference on Computational Intelligence and Data Science. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050918309335

[10] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999–7019, 2022.

[11] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18–31, 2019, doi: 10.1109/TAFFC.2017.2740923.

[12] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1195–1215, 2022, doi: 10.1109/TAFFC.2020.2981446.

[13] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, "Training deep networks for facial expression recognition with crowd-sourced label distribution," in *Proceedings of the 18th ACM international conference on multimodal interaction*, 2016, pp. 279–283.

[14] D. Kollias and S. Zafeiriou, "Expression, affect, action unit recognition: Aff-wild2, multi-task learning and arcface," *arXiv preprint arXiv:1910.04855*, 2019.

[15] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," 2020. [Online]. Available: https://arxiv.org/abs/1905.11946

[16] V.-T. Hoang and K.-H. Jo, "Practical analysis on architecture of efficientnet," in *2021 14th International Conference on Human System Interaction (HSI)*, 2021, pp. 1–4.

[17] S. Agrawal, V. Rewaskar, R. Agrawal, S. Chaudhari, Y. Patil, and N. Agrawal, "International journal of intelligent systems and applications in engineering advancements in nsfw content detection: A comprehensive review of resnet-50 based approaches," vol. 11, pp. 41–45, 10 2023.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015. [Online]. Available: https://arxiv.org/abs/1512.03385

[19] Computer Hope, "2-d," https://www.computerhope.com/jargon/num/2d.htm, 2018, accessed May 03, 2025.

[20] Fluke, "What is thermal imaging? how a thermal image is captured," https://www.fluke.com/en/learn/blog/thermal-imaging/how-infrared-cameras-work, accessed May 02, 2025.

[21] I. Liu, F. Liu, Q. Zhong, F. Ma, and S. Ni, "Your blush gives you away: detecting hidden mental states with remote photoplethysmography and thermal imaging," 2024. [Online]. Available: https://arxiv.org/abs/2401.09145

[22] V. H. Aristizabal-Tique, M. Henao-Pérez, D. C. López-Medina, R. Zambrano-Cruz, and G. Díaz-Londoño, "Facial thermal and blood perfusion patterns of human emotions: Proof-of-concept," *Journal of Thermal Biology*, vol. 112, p. 103464, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306456523000050

[23] S. S. Ajit Avasthi, Abhishek Ghosh and S. Grover, "Etichs in medical research: General principles with special reference to psychiatry research," https://pmc.ncbi.nlm.nih.gov/articles/PMC3574464/, accessed june 8, 2025.

[24] K. institutet, "Personal data in research," https://staff.ki.se/research-support/research-data-management/plan-your-research-data-management/personal-data-in-research, accessed june 8, 2025.

[25] Jamal S, Kahil Mustafa and Kamioka, Eiji, "Emotions detection scheme using facial skin temperature and heart rate variability," *MATEC Web Conf.*, vol. 277, p. 02037, 2019. [Online]. Available: https://doi.org/10.1051/matecconf/201927702037

[26] B. Ilikci, L. Chen, H. Cho, and Q. Liu, "Heat-map based emotion and face recognition from thermal images," in *2019 Computing, Communications and IoT Applications (ComComAp)*, 2019, pp. 449–453.

[27] B. Taha and D. Hatzinakos, "Emotion recognition from 2d facial expressions," in *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, 2019, pp. 1–4.

[28] S. Gupta and A. Sengupta, "Unlocking emotions through heat: Facial emotion recognition via thermal imaging," in *2023 3rd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET)*, 2023, pp. 1–5.

[29] T. Khajontantichaikun, S. Jaiyen, S. Yamsaengsung, P. Mongkolnam, and T. Chirapornchai, "Facial emotion detection for thai elderly people using yolov7," in *2023 15th International Conference on Knowledge and Smart Technology (KST)*, 2023, pp. 1–4.

[30] B. Thomas, A. Bhatt, and S. N. Singh, "Recognition of facial emotions using cnn architecture and fer2013," in *2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT)*, vol. 1, 2024, pp. 1–6.

[31] Y. Arun and G. S. Viknesh, "Leaf classification for plant recognition using efficientnet architecture," in *2022 IEEE Fourth International Conference on Advances in Electronics, Computers and Communications (ICAECC)*, 2022, pp. 1–5.

[32] H. Nguyen, K. Kotani, F. Chen, and B. Le, "A thermal facial emotion database and its analysis," in *Image and Video Technology*, R. Klette, M. Rivera, and S. Satoh, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 397–408.

[33] Q. W. S. A. S. R. S. K. R. R. S. R. e. a. Panetta, Karen, "A comprehensive database for benchmarking imaging systems," https://tdface.ece.tufts.edu/, 2018, accessed june 8, 2025.