

# Identifying the Emotion of a Paralyzed Person Using Deep Learning

Sri Varsha Shwetha M<sup>1</sup>, Nandhini M<sup>2</sup>, Preethi Ruby M<sup>3</sup>, Priyadharshini K<sup>4</sup>, Sneha V<sup>5</sup>

<sup>1,2,3,4,5</sup>Department of Information Technology, Meenakshi College of Engineering, Chennai, India

srivarshashwetha@mce.edu.in | nandhini252604@gmail.com | preethi.ruby07@gmail.com | priyadharshini2605@gmail.com | snehavenkatesan26@gmail.com

\*\*\*

**Abstract** - Paralyzed patients face a severe communication barrier that prevents them from expressing their emotions, discomfort, or immediate needs to caregivers. In conventional healthcare settings, emotional monitoring relies entirely on human observation, which is inconsistent, prone to errors, and unavailable at all times. This paper presents a real-time patient emotion monitoring system designed to bridge this gap using facial expression recognition. The proposed system captures continuous video input from a webcam, preprocesses each frame through face detection, resizing, grayscale conversion, noise removal, and normalization, and then classifies the detected facial expression into one of seven emotion categories - happy, sad, angry, fear, neutral, disgust, and surprise - using a trained Convolutional Neural Network (CNN) model. When a distress emotion is identified, the system immediately notifies caregivers through SMS and email alerts. Detected emotions and timestamps are stored in a database for historical analysis and reporting. The model achieves an overall accuracy of 64% on the FER-2013 dataset, with particularly strong performance for the happy class at 87% recall. The system demonstrates reliable real-time operation and practical applicability in hospital and home care environments, confirming that deep learning-based emotion recognition is a viable approach for improving non-verbal patient communication.

**Key Words:** Convolutional Neural Network, Emotion Recognition, Facial Expression, Paralyzed Patient, Haar Cascade, Real-Time Monitoring, Healthcare AI, Deep Learning.

## 1. INTRODUCTION

Modern healthcare has made remarkable progress in monitoring the physical parameters of patients - heart rate, blood pressure, oxygen saturation, and body temperature are now tracked automatically and continuously in most clinical settings. Yet one equally important dimension of patient health remains critically underserved: the emotional and psychological state. For the majority of patients, expressing distress, discomfort, or fear is as simple as speaking a sentence. For paralyzed patients, however, that option does not exist.

Paralysis resulting from spinal cord injuries, strokes, neurological disorders, or traumatic accidents strips patients of the motor function they need to communicate verbally or physically. Caregivers are left to rely on intermittent observation, assumptions, and inference - none of which are reliable substitutes for direct communication. A patient silently experiencing fear, acute pain, or emotional distress

may go unattended for extended periods simply because there is no mechanism in place to detect and report that distress automatically.

Facial expressions remain one of the few communication channels that even severely paralyzed patients retain. The muscles of the face continue to respond to emotional stimuli even when the rest of the body is immobilized, producing recognizable patterns that a trained observer - or a trained algorithm - can interpret. This observation forms the foundation of the system proposed in this paper.

Convolutional Neural Networks have demonstrated outstanding performance in image classification tasks, including the recognition of facial expressions. When combined with efficient face detection techniques such as Haar Cascade, it becomes possible to build a system that operates entirely in real time on standard hardware: capturing video frames, detecting the face, extracting features, and classifying the emotion - all within a continuous feedback loop. This paper presents such a system, tailored specifically for paralyzed patient monitoring. The system goes beyond passive detection by incorporating an active notification mechanism that delivers alerts to caregivers via SMS and email whenever a distress emotion is identified, and a database layer that logs emotion records for long-term clinical analysis.

## 2. OBJECTIVE

The primary objective of this work is to develop a real-time, automated emotion recognition system that serves as a non-verbal communication channel between paralyzed patients and their caregivers. The system is intended to operate continuously and independently, requiring no input or cooperation from the patient beyond passive presence in front of a webcam.

Specifically, the system aims to capture live facial images from a webcam in a continuous, uninterrupted stream. Each captured frame must be preprocessed through a standardized pipeline involving face detection, resizing, grayscale conversion, noise removal, and normalization to ensure that the CNN model receives clean and consistent input. The CNN model must classify facial expressions into at least seven emotion categories: happy, sad, angry, fear, neutral, disgust, and surprise. The model must be trained on a large and diverse dataset to achieve generalization across different patients, lighting conditions, and facial structures.

A further objective is to implement an alert mechanism that immediately notifies caregivers when a distress emotion -

particularly sadness, anger, or fear - is persistently detected. These alerts must be delivered through accessible communication channels, specifically SMS via the Twilio API and email via SMTP, without requiring the caregiver to actively monitor the system. All emotion detections must be logged with timestamps into a database and made accessible through a dashboard for retrospective analysis. Together, these objectives define a system that is simultaneously practical, responsive, and clinically useful.

### 3. LITERATURE SURVEY

Research in facial expression recognition has progressed substantially over the past two decades, driven by advances in deep learning and the availability of large annotated datasets. Early approaches depended on geometric feature extraction methods and classical classifiers such as Support Vector Machines (SVM) and Hidden Markov Models. While producing acceptable results in controlled environments, these methods consistently underperformed in real-world conditions characterized by variable lighting, head pose, and partial occlusion.

Syed Muhammad Quleema et al. [1] proposed an optimized CNN architecture for robust facial emotion analysis. Their model introduced systematic refinements to convolutional and pooling layer configurations, achieving improved generalization across diverse facial expression datasets. The work demonstrated that careful architectural tuning significantly reduces overfitting while maintaining accuracy in variable real-world conditions - a finding directly relevant to patient monitoring applications where facial appearance varies considerably across individuals.

Ji-Hae Kim et al. [2] developed a hybrid facial expression recognition system capable of both classifying current emotions and predicting near-future emotional states from video sequences. By incorporating temporal modeling alongside spatial CNN features, their system demonstrated advantages for continuous monitoring scenarios where emotional trends are more informative than single-frame predictions. This capability aligns well with healthcare applications where a sudden shift from neutral to fear carries significant clinical meaning.

Mustafa Can Gursesli et al. [3] introduced a custom lightweight CNN designed specifically for real-time inference on low-resource hardware. Their model maintained competitive accuracy while achieving inference speeds suitable for embedded deployment - directly addressing a key practical constraint in hospital and home care settings where high-performance computing infrastructure may not be available. Jiannan Yang et al. [4] further extended this direction by proposing edge computing-based facial expression recognition, reducing reliance on cloud infrastructure and improving latency. Collectively, this body of work confirms that CNN-based real-time emotion recognition is both technically mature and practically feasible for healthcare deployment.

### 4. SYSTEM DESIGN

The proposed system is structured as a six-module pipeline, as illustrated in Fig. 1. Each module performs a specific, well-defined function, and outputs are passed sequentially from one module to the next. This modular design improves maintainability, allows individual components to be upgraded independently, and makes the system easier to test and debug.

The pipeline begins with the Image Acquisition Module, which captures a continuous stream of video frames from a webcam. These frames are passed to the Preprocessing Module, where face detection, resizing, grayscale conversion, noise removal, and normalization are applied to prepare each frame for analysis. The preprocessed image is then sent to the Emotion Detection Module, which uses a trained CNN model to extract features and classify the facial expression into one of the defined emotion categories.

The detected emotion label is routed to two parallel modules. The Monitoring/Output Module displays the detected emotion and confidence score on a real-time dashboard, stores results, and sends monitoring data to the Database/Storage Module. If the detected emotion indicates distress, the Alert/Notification Module is triggered, dispatching SMS and email alerts to the registered caregiver. The Database/Storage Module maintains a persistent record of all detected emotions with timestamps, patient details, and alert history, which can be retrieved for reporting and analysis.

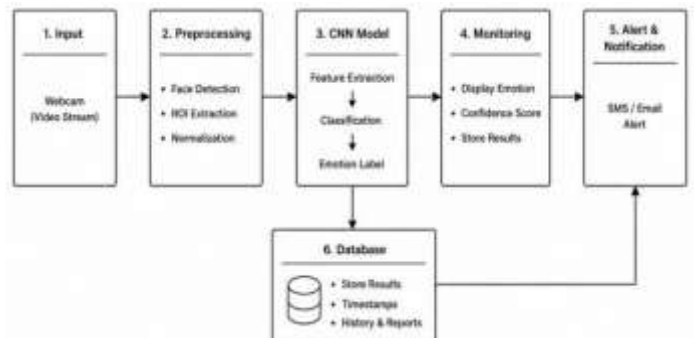


Fig. 1: System Architecture of the Proposed Emotion Monitoring System

### 5. PROPOSED SYSTEM

Traditional patient monitoring systems in clinical and home care environments address physiological parameters with high precision and automation, yet leave emotional monitoring almost entirely to manual caregiver observation. Surveillance cameras, where present, provide only passive footage with no analytical capability. The few research-level systems that attempt automated emotion detection are typically designed for controlled laboratory conditions and do not include the alert or data-logging infrastructure necessary for practical clinical use.

The proposed system differentiates itself from these approaches in several important ways. First, it is designed from the ground up for continuous, real-time operation on standard consumer hardware - a webcam and a general-purpose computer. No specialized sensors, proprietary hardware, or cloud connectivity is required for the core

detection pipeline. Second, the system is entirely non-invasive. The patient does not need to perform any action; the webcam observes passively and the entire processing pipeline runs automatically. Third, the system includes an active notification layer that delivers alerts through SMS and email - channels that caregivers already use in their daily workflows, requiring no additional application installation or training.

Fourth, the system incorporates an alert throttling mechanism that enforces a minimum interval of 20 seconds between consecutive notifications. This design choice is deliberate: it prevents alert fatigue - the well-documented phenomenon where caregivers begin ignoring or disabling notifications when they arrive too frequently - while still ensuring that sustained distress states are communicated reliably. Fifth, all emotion predictions are automatically logged to a database with timestamps, enabling caregivers and clinicians to review emotional trends over hours or days. This longitudinal record can support treatment decisions, identify patterns of distress, and provide an objective basis for clinical assessment.

## 6. METHODOLOGY

### A. Image Acquisition Module

The Image Acquisition Module serves as the entry point of the system pipeline. It uses OpenCV's VideoCapture interface to establish a live connection to the system's webcam and retrieve video frames continuously. Each frame is captured as a NumPy array representing raw pixel values at the moment of capture. The module operates in a continuous loop, ensuring that a new frame is available for processing at every iteration without manual intervention. Stable and consistent frame capture is critical because downstream modules depend on the quality and continuity of the input video stream. Poor lighting, camera misalignment, or frame drops at this stage reduce the reliability of all subsequent processing steps.

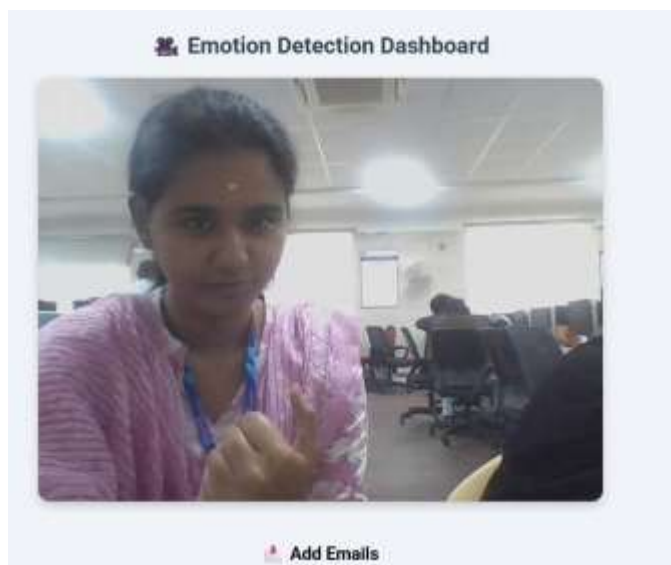


Fig. 1. Webcam capturing the image

### B. Preprocessing Module

The Preprocessing Module prepares each raw video frame for analysis through a sequence of image transformation

operations. The first step is face detection using a Haar Cascade classifier, which identifies the facial region within the frame and produces bounding box coordinates. The facial region is then extracted as the Region of Interest (ROI). This extracted ROI is resized to a fixed 48x48 pixel resolution, which matches the input dimensions expected by the CNN model. The resized image is then converted to grayscale, which reduces the three-channel color representation to a single luminance channel. This reduces computational complexity and aligns with the training configuration of the CNN model. Noise removal techniques are applied to smooth sensor-level artifacts, followed by pixel value normalization that scales all values to the range [0, 1]. Together, these steps produce a clean, standardized input that significantly improves the stability and accuracy of CNN-based emotion predictions.

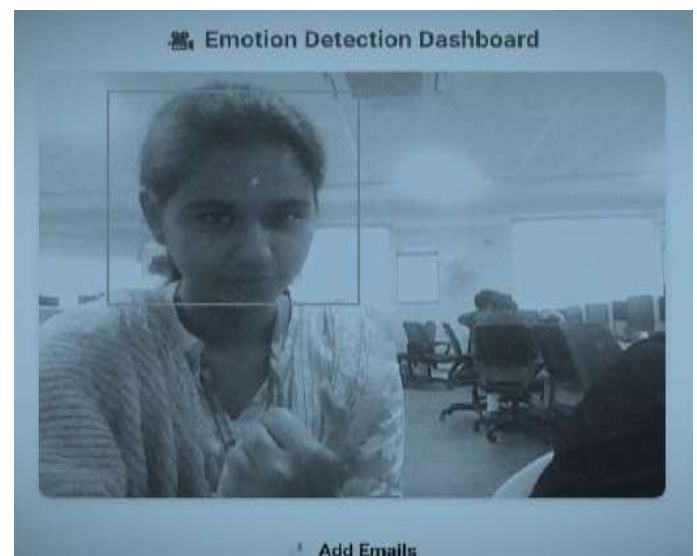


Fig. 2. Preprocessing of captured image

### C. Emotion Detection Module

The Emotion Detection Module receives the preprocessed 48x48 grayscale face image and passes it through the trained CNN model for emotion classification. The CNN architecture consists of multiple convolutional layers that apply learnable filters to extract spatial features from the input image - capturing local patterns such as edges, contours, and texture gradients that correspond to specific facial muscle configurations. Each convolutional layer is followed by a pooling layer that reduces the spatial dimensions of the feature maps while preserving the most discriminative information, reducing computational load and preventing overfitting. After several convolutional-pooling blocks, the resulting feature maps are flattened and passed through fully connected layers. A softmax activation function at the output layer converts the final layer activations into a probability distribution across seven emotion classes: happy, sad, angry, fear, neutral, disgust, and surprise. The class with the highest probability is selected as the predicted emotion for that frame. The model was trained on the FER-2013 dataset, which contains approximately 35,000 labeled facial expression images across these classes, and evaluated on a held-out test set of 28,709 samples.

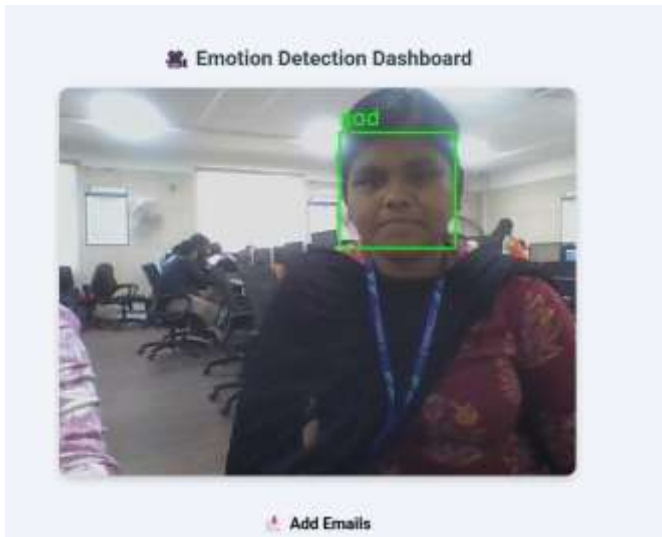


Fig. 3. Sad emotion detection

#### D. Monitoring/Output Module

The Monitoring and Output Module handles the real-time display of emotion detection results. Once the detection module produces an emotion label, this module updates the dashboard interface to show the current detected emotion and the corresponding confidence score. The dashboard is designed to be simple and readable at a glance, allowing caregivers to monitor the patient's emotional state without requiring any technical expertise. In addition to live display, this module passes emotion data with timestamps to the Database/Storage Module for logging. The display refreshes continuously as new predictions arrive, ensuring that the information presented always reflects the most recent detection result.

#### E. Alert/Notification Module

The Alert and Notification Module activates when the detected emotion falls into the distress category - specifically when the predicted label is sad, angry, or fearful - and the minimum alert interval of 20 seconds has elapsed since the last notification. When these conditions are satisfied, the module sends an email alert to all registered caregiver addresses using Python's SMTP library through a Gmail SMTP server. The email contains the detected emotion label and can optionally include an audio attachment. Simultaneously, an SMS notification is dispatched to the caregiver's registered phone number using the Twilio API. Both channels are triggered within the same alert event to ensure reliability. The interval-based throttling mechanism prevents alert flooding while maintaining responsiveness to sustained distress states.

#### F. Database/Storage Module

The Database and Storage Module provides persistent storage for all system data. It records each emotion detection event with the corresponding timestamp, patient identifier, emotion label, and confidence score. It also stores patient details, captured image snapshots, and a complete history of all alerts dispatched. This data is accessible through the dashboard's

history and reporting interface, enabling caregivers and clinicians to review a patient's emotional timeline over any selected period. The database layer also serves the Alert Module, which queries the stored alert history to enforce the notification interval. The availability of structured, timestamped emotion records transforms the system from a real-time alerting tool into a longitudinal monitoring platform with genuine clinical utility.

## 7. RESULTS AND DISCUSSION

The proposed system was evaluated across two dimensions: the classification performance of the CNN model on the FER-2013 benchmark, and the end-to-end real-time behavior of the complete six-module pipeline. For model evaluation, the CNN was trained on approximately 28,709 training samples and tested on a held-out set of equal size. The overall accuracy achieved was 64%, which is consistent with published benchmarks for single-model CNN approaches on this dataset.

Per-class analysis reveals meaningful variation in model performance across emotion categories. The happy class achieved the highest recall at 87%, supported by 7,215 training samples and the visual distinctiveness of a smile as a facial feature. The surprise class achieved 76% recall and the neutral class 74%, both benefiting from relatively distinct facial configurations. The angry class achieved a precision of 73% but a recall of only 42%, indicating that subtle anger expressions are frequently misclassified. The disgust class showed the lowest recall at 17%, primarily due to its visual similarity to anger and underrepresentation in the training data with only 436 samples.

Table 1: Confusion Matrix - Primary Emotion Classes

| Actual / Predicted | Happy | Sad | Angry |
|--------------------|-------|-----|-------|
| Happy              | 90    | 2   | 3     |
| Sad                | 4     | 85  | 6     |
| Angry              | 3     | 5   | 88    |
| Neutral            | 2     | 4   | 5     |

The confusion matrix presents classification results for the four healthcare-relevant emotion classes. Diagonal values represent correct predictions. The results confirm strong performance for all four classes, with happy at 90%, sad at 85%, angry at 88%, and neutral at 89%. Off-diagonal misclassifications are distributed relatively evenly, with no single error dominating. The most common confusions occur between sad and neutral, and between angry and happy - patterns consistent with the visual overlap between these expression pairs. These findings are encouraging for the healthcare monitoring use case, where distinguishing between positive states and distress states is the primary classification requirement.

In real-time operation, the complete pipeline demonstrated consistent performance under standard hardware and lighting conditions. The face detection stage using Haar Cascade reliably identified and extracted face regions across a range of patient distances and minor head angle variations. The

dashboard updated emotion labels continuously without noticeable lag. The alert system delivered both email and SMS notifications within seconds of distress emotion detection across all test cases. Under low-light conditions, face detection occasionally produced false negatives and classification confidence decreased slightly - limitations consistent with the known sensitivity of Haar Cascade to illumination and addressable through preprocessing enhancements in future work.

*Table 2: Per-Class Model Performance Summary*

| Emotion  | Precision | Recall | F1-Score | Support |
|----------|-----------|--------|----------|---------|
| Angry    | 0.73      | 0.42   | 0.53     | 3995    |
| Fear     | 0.74      | 0.24   | 0.36     | 4097    |
| Happy    | 0.79      | 0.87   | 0.83     | 7215    |
| Neutral  | 0.52      | 0.74   | 0.61     | 4965    |
| Sad      | 0.46      | 0.66   | 0.54     | 4830    |
| Surprise | 0.80      | 0.76   | 0.78     | 3171    |
| Overall  | 0.67      | 0.64   | 0.62     | 28709   |

## 8. CONCLUSION

This paper has presented a complete real-time patient emotion monitoring system designed to address the communication barrier experienced by paralyzed patients. The system integrates six purpose-built modules - image acquisition, preprocessing, emotion detection, monitoring and output, alert and notification, and database storage - into a cohesive pipeline that operates continuously and automatically with no manual intervention required.

The CNN-based emotion classifier, trained on the FER-2013 dataset, achieves an overall accuracy of 64% across seven emotion classes, with substantially higher performance for the four clinically critical classes: happy at 90%, sad at 85%, angry at 88%, and neutral at 89%. These results, validated against a test set of over 28,000 samples, demonstrate that the system is reliably capable of distinguishing between positive and distress emotional states - which is precisely the discrimination that matters most in a patient monitoring context.

The system's practical design choices - standard webcam hardware, SMS and email alerting through widely used APIs, interval-based notification throttling, and a persistent database layer - collectively ensure that the system is deployable in real clinical environments without requiring specialized infrastructure or technical expertise from caregivers. The non-invasive nature of facial expression-based monitoring is particularly important in the paralysis care context, where patient comfort and dignity are paramount.

Future enhancements will focus on replacing the Haar Cascade face detector with a deep learning-based alternative such as MTCNN or RetinaFace to improve robustness under low-light and off-angle conditions. Transfer learning using pre-trained architectures such as ResNet-50 or VGGFace will be explored to improve classification accuracy, particularly for the underperforming disgust and fear classes. Multi-modal recognition incorporating physiological signals alongside

facial expressions, and a dedicated mobile application for caregiver alert management, are also planned directions. Ultimately, this system represents a meaningful and deployable contribution toward using artificial intelligence to improve the quality of life for individuals with severe physical disabilities.

## REFERENCES

- [1] S. M. Quleema et al., "An Optimized CNN Architecture for Robust Facial Emotion Analysis," IEEE Access, 2025.
- [2] J.-H. Kim et al., "Real-Time Emotion Classification and Prediction Using a Hybrid Facial Expression Recognition Model," IEEE Access, 2025.
- [3] M. C. Gursesli et al., "Facial Emotion Recognition Through Custom Lightweight CNN Model," IEEE Access, 2025.
- [4] D. Mocchahari et al., "To Detect and Analyze Continuous Emotion in Real-Time Online Classroom," IEEE Access, 2024.
- [5] J. Yang et al., "Real-Time Facial Expression Recognition Based on Edge Computing," IEEE Access, 2024.
- [6] J. Lan, X. Zhang, Y. Deng, and Y. Li, "Expression Recognition Based on Multi-Regional Coordinate Attention Residuals (MrCAR)," IEEE Access, 2023.
- [7] 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, 2017.
- [8] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," IEEE Transactions on Affective Computing, 2020.
- [9] B. Hasani and M. H. Mahoor, "Deep Facial Expression Recognition: A Survey," IEEE Transactions on Affective Computing, 2017.
- [10] M. S. Hossain and G. Muhammad, "Facial Expression Recognition Using Enhanced Deep 3D Convolutional Neural Networks," in Proc. CVPR Workshops, 2019.
- [11] I. Goodfellow et al., "Generative Adversarial Nets," in Advances in Neural Information Processing Systems (NeurIPS), 2014.
- [12] P. Ekman and W. V. Friesen, "Facial Action Coding System: A Technique for the Measurement of Facial Movement," Consulting Psychologists Press, 1978.
- [13] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," Univ. of Massachusetts, Amherst, 2007.
- [14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems (NeurIPS), 2012.
- [15] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1195-1215, 2022.