REVIEW

Open Access

Facial expression deep learning algorithms in the detection of neurological disorders: a systematic review and meta-analysis



Shania Yoonesi^{1†}, Ramila Abedi Azar^{2†}, Melika Arab Bafrani^{3†}, Shayan Yaghmayee⁴, Haniye Shahavand⁵, Majid Mirmazloumi⁶, Narges Moazeni Limoudehi⁷, Mohammadreza Rahmani⁸, Saina Hasany⁹, Fatemeh Zahra Idjadi¹⁰, Mohammad Amin Aalipour¹¹, Hossein Gharedaghi¹², Sadaf Salehi¹³, Mahsa Asadi Anar^{14*} and Mohammad Saeed Soleimani^{15*}

[†]Shania Yoonesi, Ramila Abedi Azar and Melika Arab Bafrani have contributed equally to this work.

*Correspondence: Mahsa.boz@gmail.com; Soliemani.saeed@gmail.com

¹⁴ College of Medicine, University of Arizona, 1501 N Campbell Ave, Tucson, AZ 85724, USA ¹⁵ Fasa University of Medical Science, Fars, Iran Full list of author information is available at the end of the article

Abstract

Background: Neurological disorders, ranging from common conditions like Alzheimer's disease that is a progressive neurodegenerative disorder and remains the most common cause of dementia worldwide to rare disorders such as Angelman syndrome, impose a significant global health burden. Altered facial expressions are a common symptom across these disorders, potentially serving as a diagnostic indicator. Deep learning algorithms, especially convolutional neural networks (CNNs), have shown promise in detecting these facial expression changes, aiding in diagnosing and monitoring neurological conditions.

Objectives: This systematic review and meta-analysis aimed to evaluate the performance of deep learning algorithms in detecting facial expression changes for diagnosing neurological disorders.

Methods: Following PRISMA2020 guidelines, we systematically searched PubMed, Scopus, and Web of Science for studies published up to August 2024. Data from 28 studies were extracted, and the quality was assessed using the JBI checklist. A metaanalysis was performed to calculate pooled accuracy estimates. Subgroup analyses were conducted based on neurological disorders, and heterogeneity was evaluated using the *I*² statistic.

Results: The meta-analysis included 24 studies from 2019 to 2024, with neurological conditions such as dementia, Bell's palsy, ALS, and Parkinson's disease assessed. The overall pooled accuracy was 89.25% (95% CI 88.75–89.73%). High accuracy was found for dementia (99%) and Bell's palsy (93.7%), while conditions such as ALS and stroke had lower accuracy (73.2%).

Conclusions: Deep learning models, particularly CNNs, show strong potential in detecting facial expression changes for neurological disorders. However, further work is needed to standardize data sets and improve model robustness for motor-related conditions.



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicate of therwise in a credit line to the material. If material is not included in the article's Creative Commons licence, and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Keywords: Deep learning, Facial expression recognition, Neurological disorders, Metaanalysis, Convolutional neural networks

Introduction

Neurological diseases are a spectrum of conditions that may involve any neuronal tissue in our body, including the brain, spine, or peripheral neural system (PNS) [1]. Structural, chemical, and other abnormalities of nervous tissues can be the etiology of these disorders. They are classified under various categories, ranging from common disorders like Alzheimer's disease to as rare as Angelman syndrome [2, 3].

From an epidemiological aspect, neurological disorders have posed a huge burden on global health. According to GBD 2021, neurological disorders and other conditions affecting the nervous system were the leading cause of disability-adjusted life years (DALY), affecting 3.40 billion individuals [4]. Another study by Yi Huang et al. demonstrated that 805.17 million individuals suffered from neurological disorders in 2019 [5]. Some types of neurological disorders have greater significance. For instance,6 neurological conditions encompassing AD and other dementias, PD, epilepsy, MS, MND, and headache disorders ranked 9th among the leading causes of DALY [6]. Analyzing previous studies reveals that there is a significant increase in the global and territorial prevalence of neurological disease; this may be related to several factors, including increased life expectancy, improvement in socioeconomic status, advancement in screening and diagnostic devices, etc. [5, 6].

Symptoms of neurological disease vary based on the underlying cause, but some common signs and manifestations are observed across several neurological disorders. Altered facial expression is a common sign of many neurological disorders. This is due to several reasons, such as abnormalities in neuron–muscle performance [7]. For example, in Parkinson's disease, muscle stiffness leads to a characteristic symptom called a masked face [8]. In other conditions, such as a stroke, ischemia in particular brain and motor cortex regions leads to abnormalities in facial muscle control, which causes changes in facial expression [9]. Similarly, in autism spectrum disorders, abnormalities in neural processing lead to problems in the production of facial expressions [10]. In other words, it is difficult to recognize emotion based on facial expressions when there is a neurode-generative disease [11]. Based on the previous statement, it can be inferred that facial expression changes may serve as a significant indicator for a range of neurological disease. In this domain, analyzing facial expressions and improving facial emotion recognition tools may help diagnose and monitor these disorders [12–14].

Like other diseases, Effective management is one of the main goals of health care in neurological disorders. The complexity and difficulty of them make their management challenging [15]. Several factors contribute to this problem, including the nature of the disorder, diagnostic delays [16], comorbidities [17], and the need for multidisciplinary care [18]. Given these challenges, we need access to methods that facilitate screening, prediction, and even early diagnosis of neurological diseases before symptoms become apparent. Such advances lead to better prognoses and improved treatment outcomes. We can, therefore, address this issue using deep learning technology, since it has been widely adopted in recent years [19–23]. Deep learning may play a significant role in the field of neural disorders, especially in neuroimaging processes.

One of the most common models that has been used for processing medical images is the Convolutional Neural Network(CNN), which is utilized for processing medical images like MRI scans [20, 24, 25].

As we said previously, altered facial expression is a common symptom of various neurological disorders; in addition, many studies have indicated the effectiveness of deep learning/machine learning-based methods for the identification and prognosis assessment of neurological disorders. Afsaneh Davodabadi et al. indicated that artificial intelligence may be useful for diagnosis of Alzheimer's disease [11].in another study, done By Yang Wang et al., it has been demonstrated that a new CT base method utilizing DL, in terms of Cranial Automatic Planbox Imaging Towards AmeLiorating neuroscience (CAPITAL–CT), can simplify the work of radiologists and assist in neuroscience research and management of stroke [26]. Gozde Yolc et al. proposed a system consisting of two CNN steps that assist in image classification and analysis [20].

Therefore, we conduct a systematic review and meta-analysis of deep learning applications in detecting facial expression changes, aimed at diagnosis and monitoring of neurological diseases.

Methods

This systematic review follows the principles of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA2020) statement [1]. The study protocol has been registered on the Open Science Framework (OSF) (registration doi:).

Search strategy

We collected original articles in this field by searching through PubMed, Google Scholar, Web of Science, and Scopus databases for English language literature published up to the 31st of August 2024. The search was conducted based on "deep learning ", "convolutional neural network", "CNN", "DBN" and "facial expression", "FER", "Facial recognition", "neurological disorder", "neurological disease", "Alzheimer's Disease", "Acute Spinal Cord Injury", "Amyotrophic Lateral Sclerosis", "ALS", "Ataxia", "brain tumor", "Bell's Palsy", "Cerebral Aneurysm", "Epilepsy", "seizure", "Headache", "Head Injury", "Hydrocephalus", "Meningitis", "Multiple Sclerosis", "Muscular Dystrophy", "neurocutaneous syndrome", "Parkinson's Disease", "stroke", "brain attack", "cluster headache", "tension headache", "migraine headache", "migraine", "Encephalitis", "Myasthenia Gravis", "Lumbar Disk Disease", "Herniated Disk" or "Guillain–Barre Syndrome" as keywords (Table 1). The reference lists of included articles and relevant studies were manually reviewed for further relevant studies.

Data extraction

This study searched Web of Science, Scopus, and PubMed until August 2024. Studies were screened by their title and abstract using the Rayyan intelligent tool for systematic reviews. Data were extracted based on predefined variables and presented in (Table 2). Duplicated records were eliminated using EndNote ver.21.

Table 1 Curated search strategies for each database

Database	Search strategy	Results
PubMed	("deep learning"[Title/Abstract] OR "convolutional neural network"[Title/Abstract] OR "CNN"[Title/Abstract] OR "DBN"[Title/Abstract]) AND ("facial expression*"[Title/Abstract] OR "FER"[Title/Abstract] OR "Facial recognition"[Title/Abstract] OR "Facial Expression"[MeSH Terms]) AND ("neurological disorder*"[Title/Abstract] OR "neurological disease*"[Title/ Abstract] OR "Alzheimer's Disease"[Title/Abstract] OR "Acute Spinal Cord Injury"[Title/ Abstract] OR "Alzheimer's Disease"[Title/Abstract] OR "Acute Spinal Cord Injury"[Title/ Abstract] OR "Anyotrophic Lateral Sclerosis"[Title/Abstract] OR "ALS"[Title/Abstract] OR "Ataxia"[Title/Abstract] OR "brain tumor*"[Title/Abstract] OR "Bell's Palsy"[Title/ Abstract] OR "Cerebral Aneurysm"[Title/Abstract] OR "Epilepsy"[Title/Abstract] OR "seizure*"[Title/Abstract] OR "Headache"[Title/Abstract] OR "Head Injury"[Title/Abstract] OR "Hydrocephalus"[Title/Abstract] OR "Meningitis"[Title/Abstract] OR "Multiple Sclerosis"[Title/Abstract] OR "Parkinson's Disease"[Title/Abstract] OR "Multiple Sclerosis"[Title/Abstract] OR "Parkinson's Disease"[Title/Abstract] OR "stroke*"[Title/ Abstract] OR "brain attack*"[Title/Abstract] OR "ingraine headache**[Title/Abstract] OR "migraine"[Title/Abstract] OR "Encephalitis"[Title/Abstract] OR "Myasthenia Gravis"[Title/ Abstract] OR "Nervous System Diseases"[MeSH Terms] OR "Lumbar Disk Disease"[Title/ Abstract] OR "herniated Disk"[Title/Abstract] OR "Guillain-Barre Syndrome"[Title/Abstract])	32
Scopus	(TITLE-ABS-KEY("deep learning") OR TITLE-ABS-KEY("convolutional neural network") OR TITLE-ABS-KEY("CNN") OR TITLE-ABS-KEY("DBN") OR TITLE-ABS-KEY("Hierarchical Learning")) AND (TITLE-ABS-KEY("facial expression*") OR TITLE-ABS-KEY("FER") OR TITLE-ABS-KEY("Facial recognition") OR TITLE-ABS-KEY("race Expression*")) AND (TITLE-ABS-KEY("neurological disorder*") OR TITLE-ABS-KEY("neurological disease*") OR TITLE-ABS-KEY("Alzheimer's Dis- ease") OR TITLE-ABS-KEY("neurological disease*") OR TITLE-ABS-KEY("Alzheimer's Dis- ease") OR TITLE-ABS-KEY("Acute Spinal Cord Injury") OR TITLE-ABS-KEY("Anyotrophic Lateral Sclerosis") OR TITLE-ABS-KEY("ALS") OR TITLE-ABS-KEY("Ataxia") OR TITLE-ABS-KEY("brain tumor*") OR TITLE-ABS-KEY("Bell's Palsy") OR TITLE-ABS-KEY("Cerebral Aneurysm") OR TITLE- ABS-KEY("Epilepsy") OR TITLE-ABS-KEY("seizure*") OR TITLE-ABS-KEY("Headache") OR TITLE- ABS-KEY("Head Injury") OR TITLE-ABS-KEY("Hydrocephalus") OR TITLE-ABS-KEY("Meningitis") OR TITLE-ABS-KEY("Multiple Sclerosis") OR TITLE-ABS-KEY("Inucour Dystrophy") OR TITLE- ABS-KEY("neurocutaneous syndrome*") OR TITLE-ABS-KEY("Parkinson's Disease") OR TITLE- ABS-KEY("incurocutaneous syndrome*") OR TITLE-ABS-KEY("Ingraine headache*") OR TITLE-ABS-KEY("tension headache*") OR TITLE-ABS-KEY("migraine headache*") OR TITLE-ABS-KEY("migraine") OR TITLE-ABS-KEY("Ingraine headache*") OR TITLE-ABS-KEY("ingraine") OR TITLE-ABS-KEY("Incephalitis") OR TITLE-ABS-KEY("Myasthenia Gravis") OR TITLE-ABS-KEY("Lumbar Disk Disease") OR TITLE-ABS-KEY("Herniated Disk") OR TITLE-ABS-KEY("Guillain-Barre Syndrome") OR TITLE-ABS-KEY("Nervous System Diseases"))	145
WOS	((((TS = ("deep learning")) OR TS = ("convolutional neural network")) OR TS = ("CNN")) OR TS = ("DBN")) AND ((((TS = ("facial expression*")) OR TS = ("FER")) OR TS = ("Facial recognition")) OR TS = ("Face Expression*")) AND (((((((((((((((((((((((((((((((((((36

Eligibility criteria

After excluding animal-based studies, non-English articles, conference abstracts, reviews, posters, case reports, case series, and studies not related to biomarker imaging, the remaining studies were included in the review if they met the following criteria: they were human-based studies, original peer-reviewed articles, and utilized or were relevant to imaging biomarkers.

Study selection and quality assessment

Two reviewers (NW, SHY), Using the RAYYAN intelligent tool for systematic reviews, analyzed and screened titles and abstracts to identify similar papers in a blinded manner.

Author Years Country Algorithm **Facial component** Neurological description disorder David Ahmedt Landmark-based 2019 Australia Optical flow Functional neuro-Aristizabal et al. model logical disorders Region-based (FND) model Epileptic seizures (ES) Xin Liu et al. 2020 China Parallel Hierarchy Eye-brow Facial Nerve Paraly-The UK Convolutional Mouth regions sis (FNP) Saudi Arabia Neural Network Bell's palsy (PHCNN) landmark features Near-eye region Yang Zhuang USA facial weakness 2020 Histogram of Ori-Near-mouth region stroke et al. ented Gradients Bell's palsy Full face (HoG) features linear Support Vector Machine (SVM) Andrea Bandini 2021 Canada Deep learning-Oro-facial gestures, such as Amyotrophic lateral puckering lips, smiling, jaw et al. based FAN sclerosis (ALS) Traditional Post-stroke (PS) movements, and eyebrowmethods, such as raising AAM, CLM, ERT, and SDM Canada Siavash Rezaei 2021 Pairwise pain Action units (AU) involved Dementia et al. detection in the PSPI score: Contrastive AU43 (eyes closed) training AU6 (cheek raiser) Multi-task learn-AU7 (eyelid tightener) AU9 (nose wrinkler) ing AU10 (upper lip raiser) AU4 (brow lowerer) Liang-Yu Chen 2022 Taiwan Customized Facial Sadness Dementia Expression Rec-Anger 56.5% diagnosed et al. Neutrality ognition System with Alzheimer's (FERS) combined Happiness disease with Al-based 43.4% with vascular algorithms dementia Prediction models included stepwise linear regression (LR), Random forest (RF), and ensemble method (EM) The model used facial expression phenotypes to predict BPSD (Behavioral and psychological symptoms of dementia) ΕM

Table 2 Summary characteristics of included studies

Author	Years	Country	Algorithm description	Facial component	Neurological disorder
Yiming Fan et al.	2022	China	Convolutional Neural Networks (CNNs) and Vision Transformer (ViT) Pyramid Trans- former (PTF) Valence-Arousal- Like Classifier (V-ALC Lightweight model	Facial Action Coding Sys- tem (FACS) used to define expressions: Happy (AU6 + AU12) Sad (AU1 + AU4 + AU15) Angry (AU4 + AU5 + AU7 + AU23) Surprised (AU1 + AU2 + AU5 + AU26) Painful (AU4 + AU6/ AU7 + AU9/AU10 + AU43) Strained (AU4 + AU6 + AU23/AU24/ AU28) Tired (AU43 + AU54) Neutral (no specific AU)	Stroke
Jen-Cheng Hou et al.	2022	France Australia	Multi-stream deep learning model Graph convolu- tional networks (GCN) Spatio-temporal features Knowledge distil- lation technique	23 Key points detected on the face, focusing on eyebrows, eyes, nose, and mouth Facial appearance (spati- otemporal features) and facial landmarks	Epileptic seizures, particularly hyperki- netic seizures (HKN) associated with dystonia and emo- tional signs
Richard Jiang et al.	2022	UK USA UAE	Deep Learning- based framework Partial Homomor- phic Encryption (PHE) Neural Network Architecture: Deep Autoen- coder with 4 layers RMSProp as the training algorithm Edge-oriented privacy-preserv- ing system	Facial patterns Facial masking (hypomimia)	Parkinson's Disease (PD)
Davodabadi et al.	2023	Iran	Support vector machine Fractal model Collection of characteristics	Anger Disgust Fear Happy Neutral Sadness Surprise	Alzheimer's disease Mental health
Davodabadi et al.	2023	Iran	Deep convolu- tional neural net- work architecture Weight-sharing approach	Anger Disgust Fear Happy Neutral Sadness Surprise	Alzheimer's disease Mental health
Araújo de San- tana et al.	2023	Brazil	Artifcial Intelli- gence algorithms Classifying emo- tions from physi- ological signals	Hapiness Sadness Neutral Disgust Amusement Anger	Dementia

Table 2 (continued)

Author	Years	Country	Algorithm description	Facial component	Neurological disorder
Araújo de San- tana et al.	2023	Brazil	Artifcial Intelli- gence algorithms Speech signals	Neutral Calm Joy Sadness Anger Fear Surprise Disgust	Dementia
Araújo de San- tana et al.	2023	Brazil	Artifcial Intelli- gence algorithms Facial expressions	Anger Disgust Fear Happy Neutral Sad Surprise	Dementia
Huang et al.	2024	China	LeNet VGG C3D VideoMAE	Happiness Anger Surprise Disgust Sadness Fear	Parkinson's disease
Jiang et al.	2022	United States	Montreal Cogni- tive Assessment (MoCA) CNN-based facial emotion recogni- tion framework VGG16 backbone for face detection VGG19 for emo- tion classification	Neutral Happiness Sadness Surprise Fear Disgust Anger	Cognitive impair- ment Alzheimer's Disease Mild Cognitive Impairment (MCI) Vascular Dementia Frontotemporal Dementia Lewy Body Demen- tia
Shin Lim et al.	2022	Taiwan	Logistic regres- sion Random forest AdaBoost	Eye blinking (EAR) Mouth to eye distance variance Mouth height and width movement vari- ances, Peri-oral area movement variance Mouth angle variance	Parkinson's disease (PD) Motor disability Hypomimia
Oliveira et al.	2024	Brazil Australia	XGBoost Logistic regres- sion	KISS OPEN SPREAD PA PATAKA BBP BLOW	Post-Stroke condi- tions
Oliveira et al.	2023	Brazil Australia	Conditional Gen- erative Adversarial Networks (CGAN) Test-Time Aug- mentation (TTA) Logistic Regres- sion Random Forest	Smiling Disgusted Surprised	Parkinson's Disease (PD) Hypomimia
Patel et al.	2024	United States Germany Australia	ResNet-50 clas- sifier VGGFace2 HyperStyle GAN- inversion Support Vector Machine (SVM)	Smiling Non-smiling	Williams syndrome (WS) Angelman syn- drome (AS) Noonan syndrome (NS) 22q11.2 deletion syndrome (22q)

Page 8 of 25

Author	Years	Country	Algorithm description	Facial component	Neurological disorder
Singh et al.	2023	United states	convolutional neural networks (C-CNN) weakening and lesioning nodes	Masking of facial land- marks	Alzheimer's disease prosopagnosia
Singh et al.	2023	United states	Siamese networks (SN) weakening and lesioning nodes	Masking of facial land- marks	Alzheimer's disease prosopagnosia
Spahić et al.	2024	Bosnia Herzegovina Croatia Serbia	CNN	Facial movements	neurological impair- ment disorder cerebral palsy pilepsy autism spectrum disorder
Tanaka et al.	2022	Japan	ASEA (Assess- ment Scale for Engagement in Activities) Kokoro sensor	Anger Contempt Disgust Engagement Fear Joy Sadness Surprise Valence	dementia Alzheimer's Disease (AD) Vascular Dementia (VaD) Dementia with Lewy Bodies (DLB)
Chandra Tripathi et al.	2024	India	Active Appear- ance Models (AAM) Consistent response meas- urements (CRM) Facial Action Coding System (FACS) Keypoint Move- ment (KPM) vector Statistical mod- eling of KPM	Happy Surprise Fear Disgust Sadness	Stroke ADHD Alzheimer's Parkinson's Schizo- phrenia Autism Depression Epileptic Seizures Dementia
Wang et al.	2024	China	Convolutional neural network (CNN) Age and sex matched samples Data augmen- tation-Image flipping Novel structure networks Fine tuning with pretrained weights Model ensemble	Changes in the central facial area Distances between the lower face contour Drooping eyes and mouth corners	Acute Ischemic Stroke
Wang et al.	2022	China	CAPITAL-CT V-Net Region proposal network (RPN)	Facial boundaries	Stroke
Zhu et al.	2022	China	Multiview con- volutional neural network	Facial EMG patterns	Peripheral facial nerve palsy

Full texts of them were obtained to assess the qualification of the "Yes" and "Maybe" groups. In case of conflicts, a third reviewer (MAA)was involved and then reached an agreement to overcome differences and disagreements. Conflicts have been resolved through discussion between them. Quality assessment of each article was performed independently by two authors using the JBI checklist. For each included study, the quality assessment and risk of bias were performed and evaluated using JBI's critical appraisal tools.

Statistical analysis

A meta-analysis was performed to calculate pooled accuracy estimates for evaluating the performance of deep learning algorithms in facial expression recognition for detecting neurological disorders. The heterogeneity among studies was assessed using the Chi-square test and quantified by the I^2 statistic, which indicates the proportion of variability due to heterogeneity rather than random fluctuation. The I^2 statistic was computed with the formula: $100\% \times (Q - df)/Q$. Study weights were determined using the inverse variance method. A random-effects model was applied to combine data from the studies, reducing the impact of heterogeneity. Statistical significance was set at P < 0.05. Data from graphical figures in the studies were digitized using WebPlot Digitizer (Automeris LLC, Frisco, TX, USA).

Regression models and imputation techniques were employed to address missing accuracy values in the meta-analysis. These methods were utilized to handle missing data systematically, ensuring a more accurate synthesis of the study results. Regression models predicted missing values based on the relationships observed in the data, while imputation methods (e.g., multiple imputation) replaced missing values with plausible estimates, considering the uncertainty inherent in the imputation process. These approaches were selected to reduce bias and enhance the accuracy of the pooled estimates, ensuring the meta-analysis results offer a reliable representation of the underlying data. In addition, these methods help preserve the statistical power of the analysis by incorporating all available data rather than excluding studies with missing values, which could lead to bias or affect the generalizability of the findings. For future analyses, methods such as maximum likelihood estimation and Bayesian techniques could also be explored to more effectively handle missing data.

Results

Study characteristics

This systematic review and meta-analysis included 28 studies published between 2019 and 2024, investigating the application of deep learning-based facial expression analysis for detecting neurological disorders (Fig. 1). The studies were conducted across multiple countries, including Australia, China, the United Kingdom, Saudi Arabia, the United States, and Canada, with test sample sizes ranging from 26 to 437 images or video recordings (Table 2).

The neurological conditions assessed in the included studies encompassed dementia, Bell's palsy, stroke, amyotrophic lateral sclerosis (ALS), functional neurological disorders (FND), and motor impairments. While demographic data were inconsistently reported, studies that included age distributions showed mean participant ages ranging from 63.2



Fig. 1 Prisma flow chart of the study selection procedure

to 82.5 years. Certain studies stratified participants by neurological condition, such as Andrea Bandini et al., who reported mean ages of 63.2 years (healthy control), 61.5 years (ALS), and 64.7 years (post-stroke). However, several studies lacked explicit reporting of age, sex, and training sample characteristics, which may impact the generalizability of findings.

The studies employed a range of deep learning methodologies, including landmarkbased models, convolutional neural networks (CNNs), facial action unit (AU) analysis, and hybrid approaches. Facial components analyzed included eyebrow and mouth regions, full-face expressions, near-eye regions, and action unit-based facial dynamics.

Algorithm performance and accuracy

The classification performance of the deep learning models varied significantly across studies. Reported accuracies ranged from 1.8% to 94.8%, with some studies reporting

alternative performance metrics, such as Pearson correlation coefficients, rather than conventional accuracy rates.

- The highest classification accuracy (94.8%) was observed in Xin Liu et al. for Bell's palsy detection, using a parallel hierarchical convolutional neural network (CNN) that leveraged eyebrow and mouth region features.
- The lowest classification accuracy (1.8%) was reported by Andrea Bandini et al. for motor impairments (ALS, post-stroke symptoms), suggesting the challenges associated with distinguishing subtle facial expression alterations in these conditions.
- Studies evaluating dementia detection reported Pearson correlation coefficients rather than traditional accuracy metrics, making direct comparisons challenging. However, deep learning models trained on dementia-related facial expression data demonstrated strong predictive capabilities.
- Functional neurological disorders (FND) and stroke detection showed moderate classification accuracy, influenced by the specific facial regions analyzed and the complexity of the algorithmic approach.

Meta-analysis findings

A random-effects meta-analysis was conducted to assess the pooled accuracy of deep learning models across all included studies. The overall estimated accuracy was 89.25% (95% CI 88.75–89.73%), indicating a high predictive capability of deep learning-based facial expression analysis for neurological disorder detection (Fig. 2).

Subgroup analysis

A subgroup analysis based on neurological condition categories revealed significant variations in classification performance (Fig. 3):

- Dementia: 77.49% (95% CI 64.71-90.28%).
- Bell's Palsy: 94.58% (95% CI 91.02–98.13%).
- Stroke: 90.4% (95% CI 73.23–107.65%).
- Motor impairments (ALS, post-stroke symptoms): 73.2% (95% CI 45.82–100.59%).

These results indicate that deep learning algorithms exhibit high classification performance for dementia and Bell's palsy, while performance for motor-related conditions remains comparatively lower.

Heterogeneity and statistical analysis

The heterogeneity analysis indicated substantial variability in the reported effect sizes across studies, with an I^2 statistic of 99.11%, confirming high statistical heterogeneity. The Q test (Q=1594.24, p<0.0001) further supported the presence of significant variation among included studies, suggesting that differences in algorithm types, data set sizes, and neurological disorder classifications contributed to the observed heterogeneity.



Fig. 2 Forest plot of pooled accuracy estimates of DL models

Publication bias and sensitivity analysis

A funnel plot analysis (Fig. 4) demonstrated that most studies fell within the pseudo-95% confidence interval, indicating a low likelihood of publication bias. However, a minor asymmetry was observed, suggesting the potential influence of selective reporting.

The Galbraith plot (Fig. 5) further confirmed that while most studies aligned along the regression line, a subset exhibited outlier behavior, suggesting inconsistencies in reported classification accuracies.

Risk of bias and quality assessment

The Cochrane Risk of Bias Tool was used to assess the methodological quality of the included studies. While most studies demonstrated low risk concerning participant selection and outcome assessment, some exhibited unclear or high risk in domains, such as allocation concealment and blinding procedures. In addition, studies with small test sample sizes and incomplete demographic reporting were classified as having a higher risk of bias, potentially affecting the generalizability of findings.

	Random-effects REM Effect size	1L model Weight
Study	with 95% CI	(%)
Dementia		
Araújo de Santana et al.	99.00 [94.49, 103.51]	4.07
Araújo de Santana et al.	80.00 [75.49, 84.51]	4.07
Araújo de Santana et al.		3.98
Jiang et al.	63.30 [58.79, 67.81]	4.07
Singh et al.	85.00 [32.20, 37.74]	4.10
Heterogeneity: $r^2 = 250.02$ $l^2 = 98.58\%$ $H^2 = 70.31$		4.11
Test of $\theta_1 = \theta_1$: Q(5) = 417.74, p = 0.00		
Test of θ = 0: z = 11.88, p = 0.00		
Bell's palsy		
Zhu et al.	93.70 [85.27, 102.13]	3.92
Heterogeneity: $\tau^2 = 0.00$, $I^2 = .\%$, $H^2 = .$	93.70 [85.27, 102.13]	
Test of $\theta_i = \theta_j$: Q(0) = 0.00, p = .		
Test of θ = 0: z = 21.79, p = 0.00		
Rell'snalsv		
Xin Liu et al		3.98
Yang Zhuang et al.	94,50 [90,38, 98,62]	4.08
Heterogeneity: $r^2 = 0.00$, $l^2 = 0.00\%$, $H^2 = 1.00$	94.58 [91.02, 98.13]	
Test of $\theta_i = \theta_i$: Q(1) = 0.01, p = 0.94		
Test of θ = 0: z = 52.14, p = 0.00		
Dementia		
Siavash Rezaei et al.	48.00 [35.46, 60.54]	3.69
Liang-Yu Chen et al.	86.00 [76.59, 95.41]	3.87
Davodabadi et al.	98.80 [96.45, 101.15]	4.11
Heterogeneity: $T = 665.44$, $T = 97.68\%$, $H = 43.04$	78.08 [48.42, 107.73]	
Test of $\theta_i = \theta_j$, $Q(2) = 65.66$, $p = 0.00$		
rest 01 0 = 0. 2 = 0.10, p = 0.00		
Functional neurological disorders		
David Ahmedt Aristizabal et al.	68.10 [62.81, 73.39]	4.04
Jen-Cheng Hou et al.		3.78
Spahić et al.	93.60 [80.27, 106.93]	3.64
Chandra Tripathi et al.	78.00 [75.65, 80.35]	4.11
Heterogeneity: τ^2 = 77.00, I^2 = 89.07%, H^2 = 9.15	78.88 [69.31, 88.45]	
Test of $\theta_i = \theta_j$: Q(3) = 18.03, p = 0.00		
Test of θ = 0: z = 16.16, p = 0.00		
Straka		
Yiming Ean et al	99 80 [95 68 103 92]	4.08
Wang et al.	73.00 [71.43, 74.57]	4.12
Wang et al.	98.70 [96.35, 101.05]	4.11
Heterogeneity: r ² = 229.22, l ² = 99.33%, H ² = 148.38	90.44 [73.23, 107.65]	
Test of θ _i = θ _j : Q(2) = 391.09, p = 0.00		
Test of θ = 0: z = 10.30, p = 0.00		
genetic conditions		
Patel et al.	77.10 [72.59, 81.61]	4.07
Heterogeneity: $\tau^2 = 0.00$, $I^2 = .\%$, $H^2 = .$	77.10 [72.59, 81.61]	
Test of $\theta_i = \theta_j$: Q(0) = 0.00, p = .		
rest or $\theta = 0$: $z = 33.52$, $p = 0.00$		
motor condition		
Andrea Bandini et al.	18.00 [11.34. 24 66]	3.99
Richard Jiang et al.		3.93
Huang et al.	88.00 [84.86, 91.14]	4.10
C. Oliveira et al.	82.00 [80.39, 83.61]	4.12
C. Oliveira et al.	83.00 [74.18, 91.82]	3.90
Heterogeneity: τ^2 = 965.64, I^2 = 99.55%, H^2 = 223.60	73.20 [45.82, 100.59]	
Test of $\theta_i = \theta_j$: Q(4) = 375.22, p = 0.00		
Test of θ = 0: z = 5.24, p = 0.00		
Overall	80.71 [73.34, 88.09]	
Heterogeneity: r ² = 343.17, l ² = 99.11%, H ² = 111.81		
Test of $\theta_i = \theta_j$: Q(24) = 1594.24, p = 0.00		
lest of θ = 0: z = 21.44, p = 0.00		
Test of group differences: $Q_b(7) = 44.73$, p = 0.00	50 100	

Fig. 3 Subgroup analysis of pooled Accuracy based on neurological disorder type between DI models



Fig. 4 Funnel plot demonstrating a symmetric view confirms the lack of publication bias



Fig. 5 Galbraith plot for heterogeneity assessment

Discussion

This systematic review and meta-analysis evaluated the application of deep learning (DL) models, particularly convolutional neural networks (CNNs), in recognizing facial expression changes associated with various neurological disorders. The pooled accuracy of 89.25% across 28 studies underscores the potential of facial expression analysis as a diagnostic aid. However, significant variability in performance was observed depending on the type of neurological condition, the complexity of facial alterations, and the deep learning architecture used.

Facial expression in Parkinson's disease

Studies investigating Parkinson's disease (PD) consistently highlighted the utility of deep learning (DL) for detecting hypomimia and evaluating treatment response. Jiang et al. achieved over 95% accuracy using a model to distinguish pre- and post-DBS treatment expressions, outperforming EEG-based methods [14]. In contrast, Huang et al.'s MARNet model achieved 88% accuracy across PD stages by analyzing six core emotions, demonstrating robustness for staging purposes [27]. Models combining facial data with voice biomarkers, such as the smartphone-based approach by Lim et al., achieved an AUROC of 0.90—suggesting that multimodal inputs enhance diagnostic precision [28]. Furthermore, Oliveira et al. boosted classification performance from 62 to 83% using Conditional Generative Adversarial Networks (CGAN) for synthetic data generation and test-time augmentation for model stability [9, 29]. Collectively, while all models showed promise, those integrating multimodal data or data augmentation strategies offered greater generalizability and resilience across PD symptom presentations.

Facial expression in stroke

Stroke-related studies demonstrated a wide range of performances based on data set characteristics and model complexity. Fan et al's FER–PCVT model achieved near-perfect accuracy (99.81%) on a private data set but dropped to 88.2% and 89.4% on public data sets, revealing potential overfitting to task-specific expressions [30]. Bandini et al. emphasized the importance of domain-specific fine-tuning; their model initially underperformed in post-stroke patients until calibrated, after which normalized RMSE improved significantly [31]. In addition, Wang et al's CAPITAL–CT model provided rapid, low-dose diagnostic imaging with over 98% accuracy using CNN segmentation for facial asymmetry detection, highlighting the promise of medical image-based DL in real-time stroke diagnostics [26, 32]. These findings suggest that while CNNs and transformer hybrids can effectively classify stroke-related facial impairments, model calibration to clinical contexts and data set diversity is essential for optimal performance.

Other neurological conditions

In disorders, such as facial nerve palsy, muscular dystrophy, and general facial asymmetry, performance varied based on model architecture and the visibility of facial cues. Zhuang et al. achieved up to 89.7% accuracy for detecting facial weakness using a land-mark-based model combined with HoG features and SVM classifiers [33]. Tripathi et al. introduced a metric that correlated with manual FACS scoring (R^2 =0.78), validating its use for identifying asymmetry in stroke or ALS [34]. Zhu et al. improved diagnostic performance by combining surface electromyography (sEMG) with multiview CNNs, which also supported treatment evaluation in Bell's palsy [35]. These results highlight the importance of combining facial features with biosignal-based or task-specific cues to improve DL performance in less overt neurological presentations.

Facial expression in epileptic seizures

Research on epileptic seizures using facial expression analysis is emerging but promising. Ahmedt-Aristizabal et al. demonstrated that region-based models outperformed landmark-based approaches (79.6% vs. 68.1%) for differentiating epileptic seizures from functional neurological disorder (FND), due to better recognition of complex motion patterns [11]. Hou et al. confirmed this through a dual-model system using adaptive graph convolutional and temporal convolutional networks (AGCN+TCN), achieving~80% accuracy in detecting dystonia and emotional expressions during hyperkinetic seizures [36]. These findings suggest that video-based temporal models capturing sub-tle dynamic changes are particularly suited for seizure-related applications, although broader data set validation is still needed.

Facial expression in dementia

Dementia-related studies reinforced the diagnostic and monitoring potential of emotion recognition. Chen et al. developed the Facial Expression Recognition System (FERS), which predicted behavioral and psychological symptoms of dementia (BPSDs) with high accuracy (r=0.834), especially for negative emotional states [37]. Jiang et al. used CNNs to differentiate cognitively impaired (CI) from unimpaired individuals with 90.1% accuracy in structured environments, and improved classification when eye-tracking data was added [14]. Rezaei et al. also addressed pain expression in dementia, showing that deep learning models could identify subtle facial cues of discomfort in non-verbal populations [31]. Overall, these studies demonstrate the effectiveness of integrating facial analytics into emotion-based monitoring systems for dementia care.

Facial expression in Alzheimer's disease and other neurological conditions

Facial expression recognition has shown considerable promise in Alzheimer's disease (AD) as well as in other neurological and genetic conditions. In the context of AD, emotion classification and cognitive monitoring through facial cues were frequently explored. Davodabadi et al. compared support vector machines (SVM) and deep convolutional neural networks (DCNN), finding superior performance with the DCNN model (testing accuracy: 98.8%), particularly for classifying core emotions relevant to dementia progression [11]. Singh et al. extended this line of work by modeling AD-associated face recognition deficits using Siamese networks, demonstrating the resilience of this architecture to simulated neural degradation and drawing parallels with cognitive decline patterns in AD patients [38].

Chen et al.'s ensemble approach predicted behavioral and psychological symptoms of dementia (BPSDs) in Alzheimer's patients by analyzing emotional variability, achieving strong correlation with neuropsychiatric inventory scores (r=0.834) [37].Complementing this, Tripathi et al. developed a Consistent Response Measurement (CRM) framework to detect emotional engagement via facial key points during emotionally evocative stimuli, providing a low-cost and scalable method for early cognitive screening [34].

Beyond AD, deep learning was applied to detect facial markers in genetic and neurodevelopmental conditions. Patel et al. demonstrated that facial expressions significantly impact diagnostic performance in syndromes, such as Williams and Angelman, where accuracy dropped when typical expressions (like smiling in WS) were altered—underscoring the diagnostic weight of facial affect [39]. Similarly, Spahić et al. introduced the TRUEAID system, which used 4D ultrasound to analyze fetal neurobehavioral movements, achieving 93.8% accuracy in detecting early neurological disorders, such as autism, cerebral palsy, and epilepsy [40].

These diverse applications suggest that facial expression analysis via DL is not only valuable in neurodegenerative diseases like AD, but also holds diagnostic and prognostic utility across developmental, genetic, and pediatric neurological conditions, especially when paired with multimodal data streams or expression-specific model training.

The application of multimodal cascade transformers in facial expression algorithms

The application of multimodal cascade transformers in the context of facial expression recognition for neurological disorders holds great promise for enhancing diagnostic accuracy and robustness. In recent years, transformers have shown significant potential in handling sequential data, such as time-series information and multimodal inputs, making them highly suitable for complex tasks like analyzing facial expressions and other clinical data [41–43].

Multimodal cascade transformers are designed to process multiple types of data (e.g., facial expression videos, speech signals, and physiological data) in a cascading manner, allowing for better integration and contextual understanding. This approach could offer several advantages in detecting neurological disorders, where facial expressions alone may not provide sufficient information due to the subtlety of changes in certain conditions or the complexity of interpreting these expressions in isolation [44].

For example, a cascade structure allows the model to first process facial expressions using a dedicated transformer block, followed by the integration of complementary modalities, such as speech, physiological data (e.g., heart rate and skin conductance), and neuroimaging data in subsequent stages. Each modality can be processed separately by specialized networks, with their outputs then combined through a series of transformer layers to capture cross-modal interactions effectively. This combination could improve diagnostic performance in disorders like Parkinson's disease, where facial expressions (such as hypomimia) and speech patterns (e.g., hypophonia) are often closely linked [45].

Moreover, cascade transformers can help address neurological disorders' temporal dynamics. For example, deep learning models could analyze how facial expressions evolve over time while accounting for other signs of neurological decline from video or audio data. This multimodal processing can also lead to more accurate disease progression and treatment outcomes.

By incorporating contextual information from multiple sources, multimodal cascade transformers could reduce the limitations posed by relying on a single modality, such as facial expression alone. This makes them well-suited for applications in clinical settings, where complex, multimodal data are often available. It could lead to more holistic, reliable, and personalized diagnostic tools for neurological disorders [46].

The use of multimodal cascade transformers represents an exciting frontier in the integration of deep learning techniques for neurological disorder diagnosis. By leveraging the synergy between different data modalities, these models could significantly improve the accuracy and robustness of facial expression recognition systems, enabling better monitoring, diagnosis, and treatment planning for neurological patients [47].

Comparison with existing reviews

Several prior reviews have addressed deep learning applications in neurology, yet few have focused specifically on facial expression recognition (FER) across a broad spectrum

of neurological disorders. Valliani et al. offered a general overview of deep learning in neurology, highlighting use cases such as neuroimaging and EEG analysis but did not delve into facial dynamics or expression-based biomarkers [48]. Similarly, Gautam and Sharma conducted a meta-analysis on deep learning in neurological disorder detection, yet it centered predominantly on imaging and lacked an analysis of behavioral markers like facial expression [49].

Yolcu et al. presented a CNN-based FER system for monitoring neurological disorders but limited their focus to architecture design without performing a meta-analysis or comparing performance across conditions [12]. More recently, Alzubaidi et al. and Sajjad et al. compiled comprehensive reviews on facial expression recognition techniques, data sets, and CNN variants, but neither study explicitly focused on neurological disorder populations or provided disease-specific accuracy benchmarks [50, 51].

What distinguishes the current review is its systematic integration of 28 studies involving FER applications across both common (e.g., Alzheimer's disease and Parkinson's disease) and rare (e.g., Angelman syndrome and Williams syndrome) neurological conditions. This study also provides:

- The first meta-analysis of accuracy rates for FER models in this domain, yielding a
 pooled accuracy of 89.25% with clear subgroup estimates by disease.
- A comparative synthesis of multimodal models, highlighting where integration with speech, eye-tracking, or physiological signals (e.g., Araújo de Santana et al. [15]) leads to diagnostic improvements.
- A structured breakdown of model performance by disorder type and architectural approach (e.g., CNN, transformer, and CGAN), something not found in prior narrative reviews.

Moreover, the current review uniquely examines the clinical translational value of FER-based AI tools, particularly their utility in early detection, remote monitoring, and multidisciplinary neurological care. In doing so, it addresses the growing demand for non-invasive, low-cost diagnostic adjuncts—a focus not yet emphasized in most prior literature.

Clinical implications

The clinical implications of using deep learning algorithms, specifically convolutional neural networks (CNNs) and multimodal models like cascade transformers, in facial expression recognition for detecting neurological disorders, are profound and far-reaching. These models can transform diagnostic practices, offering more efficient, accurate, and non-invasive methods for assessing neurological health [52].

Early detection and diagnosis:

Neurological disorders often present with subtle symptoms that may not be immediately evident in traditional clinical assessments. Deep learning-based facial expression recognition systems can detect early signs of conditions such as Parkinson's disease, Alzheimer's disease, and stroke by identifying minute facial expression changes that may go unnoticed by clinicians. The ability to diagnose these disorders earlier can lead to earlier interventions, which are crucial for improving patient outcomes, slowing disease progression, and enhancing quality of life. For example, the recognition of hypomimia (masked face) in Parkinson's disease could aid in diagnosing the condition before more pronounced motor symptoms appear [50].

Non-invasive monitoring

Traditional diagnostic techniques for neurological disorders, such as neuroimaging or extensive motor assessments, can be invasive, costly, and time-consuming. Facial expression recognition offers a non-invasive alternative that can be used as a complementary diagnostic tool. Using video recordings or real-time monitoring, these deep learning systems can provide continuous, low-cost monitoring of patients. This is particularly useful for tracking disease progression in conditions, such as dementia or ALS, where regular assessment is needed to adjust treatment plans and manage symptoms effectively [51].

Real-time clinical decision support

In clinical settings, deep learning models could be integrated into decision-support systems that assist healthcare providers by analyzing facial expressions in real-time during consultations. For instance, a deep learning algorithm could analyze a patient's facial expressions during a routine check-up and provide insights into potential neurological conditions, allowing for more timely interventions. This integration could reduce diagnostic delays, especially in settings with limited access to specialist care, improving patient outcomes through earlier and more accurate assessments [53].

Personalized treatment plans

The use of deep learning in facial expression analysis can enhance personalized medicine. By integrating facial expression data with other clinical information (e.g., medical history and cognitive tests), algorithms can help create tailored treatment plans that are specific to each patient's condition and disease stage. This personalized approach could optimize therapeutic strategies, ensuring that patients receive the most effective interventions based on their unique presentation of symptoms.

Improving accessibility and reducing healthcare disparities

Implementing deep learning models in clinical practice can also help reduce healthcare disparities by making advanced diagnostic tools more accessible to underserved populations. In remote or resource-limited settings, facial expression recognition systems can provide affordable and effective screening tools without the need for expensive equipment or specialist expertise. For example, in rural areas, patients could use mobile devices equipped with facial recognition software to undergo preliminary assessments, with results analyzed remotely by healthcare professionals [54].

Aiding in multidisciplinary care

Neurological disorders often require multidisciplinary care, including neurologists, speech therapists, psychologists, and rehabilitation specialists. Facial expression recognition tools could serve as a valuable bridge for coordinating care by providing objective data on facial dynamics and helping clinicians monitor patients' emotional well-being, cognitive function, and motor control over time. This data could be shared across the care team to improve communication and ensure that all aspects of the patient's condition are addressed holistically.

Facilitating clinical trials

The integration of facial expression recognition into clinical trials could improve the monitoring of patient responses to experimental treatments. By tracking changes in facial expressions over time, researchers could gain objective insights into how patients are responding to interventions, especially in conditions, where self-reporting is unreliable (e.g., dementia). This can lead to more robust, real-time data collection during trials, which is essential for evaluating new therapies' efficacy.

In conclusion, the application of deep learning algorithms for facial expression recognition in neurological disorders offers a range of clinical benefits. These technologies have the potential to revolutionize diagnostic practices, offering earlier, more accurate, and non-invasive detection, enhancing patient care, and improving treatment outcomes. As technology advances and becomes more integrated into clinical workflows, it could lead to more efficient, accessible, and personalized care for patients with neurological conditions [55].

Limitations

While this study provides valuable insights into the potential of deep learning algorithms for detecting facial expression changes associated with neurological disorders, there are several limitations. First, the study only included peer-reviewed, open-access articles in English, which may have excluded relevant research published in other languages or databases. In addition, the studies included in this meta-analysis varied widely in terms of the neurological conditions studied, sample sizes, and methodologies, which introduced substantial heterogeneity. The lack of standardized data sets and inconsistent reporting of participant demographics, including age and sex distributions, further limited the generalizability of the findings. The high heterogeneity observed across the included studies in this meta-analysis is a notable challenge that affects the interpretation of the results. This variability stems from differences in study methodologies, such as variations in deep learning models used (e.g., CNNs and hybrid approaches), the specific neurological disorders studied, and the facial components analyzed (e.g., full-face expressions vs. specific facial regions, such as the eyes or mouth). In addition, discrepancies in sample sizes, participant demographics, and diagnostic criteria further contribute to this heterogeneity. While the random-effects model and I^2 statistic were used to quantify and mitigate the impact of this heterogeneity, the substantial variation in study designs and outcomes remains a limitation. To address this issue, future studies should aim to standardize the data collection process, including the use of consistent deep learning models, facial expression recognition protocols, and participant characteristics, which will enhance the reliability and comparability of findings across different studies. Moreover, the use of alternative performance metrics (such as Pearson correlation coefficients) instead of traditional accuracy rates in some studies made direct comparisons challenging. Finally, while multiple imputation methods were used to handle missing data, potential biases due to incomplete or missing information may persist, affecting the accuracy of the pooled estimates.

Another critical limitation not yet widely addressed in this field is the privacy and security of facial expression data, particularly in clinical applications involving sensitive neurological conditions. While a few studies such as Jiang et al. have explored privacy-preserving frameworks using homomorphic encryption for Parkinson's disease monitoring [14], most deep learning models in facial expression recognition lack built-in privacy safeguards. This gap is concerning given the sensitive nature of facial data and the ethical implications of real-time patient monitoring. Moreover, data compression and optimization strategies—such as model pruning, quantization, and knowledge distillation—remain underutilized in the reviewed literature, despite their potential to make FER models more efficient and deployable in low-resource or mobile clinical settings.

The recent review by Jafari et al. [56]highlights the increasing relevance of lightweight, privacy-aware transfer learning models in biomedical contexts. Incorporating these methods into FER systems could improve model generalizability, reduce computational cost, and mitigate data-sharing risks in multicenter research and telehealth deployment.

Future directions

Future research should focus on developing standardized data sets encompassing a wide range of neurological disorders to ensure better comparability across studies. In addition, future studies should explore the integration of multimodal data (e.g., speech, physiological signals, and neuroimaging) with facial expression recognition models to enhance diagnostic accuracy and robustness. More efforts are needed to refine deep learning models, improve their generalizability across different patient populations, and reduce biases related to demographic variations. In addition, exploring the application of advanced techniques such as Bayesian models or maximum likelihood estimation could improve the handling of missing data, thereby increasing the reliability of conclusions drawn from these models. Finally, as the field progresses, incorporating real-time facial expression recognition tools in clinical settings could provide more timely and accurate assessments of neurological conditions, offering significant benefits for patient care.

A critical challenge in deploying deep learning-based facial expression recognition (FER) systems in clinical neurology lies in ensuring data privacy and security. The sensitive nature of facial data necessitates robust privacy-preserving mechanisms, yet only a limited number of studies have explored this. For instance, Zhang et al. developed a wearable, depth-sensing FER device that analyzes facial skin deformation without capturing identifiable features, thereby maintaining patient anonymity and allowing ondevice inference with no data transmission to external servers [57]. Similarly, Jiang et al. introduced a privacy-aware Parkinson's diagnostic system employing partial homomorphic encryption to perform facial analysis directly on encrypted data, showcasing the potential of edge computing in safeguarding patient confidentiality [14].

In addition to privacy, optimization for real-world implementation remains underexplored. Many high-performing FER models require significant computational resources, posing barriers in low-resource settings or mobile deployments. Li et al. reviewed model compression techniques, such as pruning, quantization, and knowledge distillation, which can reduce model size and improve efficiency with minimal performance trade-offs [58]. Likewise, Surianarayanan et al. emphasized the role of lightweight neural architectures and memory-efficient processing for edge-AI applications like FER in clinical monitoring. These strategies are essential for scaling FER tools into routine clinical use without sacrificing speed or accessibility [58, 59].

Furthermore, federated learning presents a compelling approach for enabling collaborative model training across institutions without compromising patient privacy. Rieke et al. outlined how federated learning can facilitate decentralized data utilization, making it particularly well-suited for sensitive healthcare applications [60]. Building on this, Zhuang et al. discussed combining foundation models with federated architectures to manage data heterogeneity and optimize communication efficiency. Applying such frameworks to FER could accelerate progress in building generalizable, privacy-conscious diagnostic systems [61].

Together, these advancements underscore the need for future FER research to move beyond accuracy benchmarks and prioritize ethical deployment, computational efficiency, and cross-institutional scalability.

Conclusion

This systematic review and meta-analysis highlights the considerable promise of deep learning (DL), particularly convolutional neural networks (CNNs), in detecting facial expression changes associated with a wide spectrum of neurological disorders. The pooled accuracy of 89.25% and consistent high performance in disorders such as dementia and Bell's palsy affirm that facial expression recognition (FER) can serve as a valuable non-invasive biomarker for neurological screening, diagnosis, and monitoring.

The clinical implications of these findings are substantial. FER systems powered by DL offer a scalable, low-cost alternative to traditional neurological assessments, with potential applications in telemedicine, continuous monitoring, and early detection—particularly in resource-limited or underserved settings. Moreover, integrating FER with multimodal data (e.g., speech, eye tracking, or physiological signals) and optimization techniques like quantization or federated learning can further enhance model performance, privacy, and deployment feasibility.

By mapping model performance across different neurological conditions and identifying key architectural and methodological drivers of success, this study not only synthesizes current capabilities but also lays a foundation for future research and innovation. As the field progresses, DL-based FER tools are poised to become integral to next-generation neurological diagnostics—offering clinicians objective, interpretable, and real-time insights into complex neurocognitive and neuromotor disorders.

Author contributions

Study concept and design: MAA,MSS acquisition of the data: HS,SH analysis and interpretation of the data: MAA drafting of the manuscript: MR,NMK,SY,MM,RAA,MAA,HG,SS,FZI,MAB,MSS critical revision of the manuscript for important intellectual content: MAA administrative, technical, and material support: MAA study supervision: MAA.

Funding

No institution supported or funded this study.

Data availability

The data is available upon reasonable request from the corresponding author.

Declarations

Ethics approval and consent to participate Not applicable.

Informed consent Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Psychology, Central Tehran Branch, Islamic Azad University, Tehran, Iran. ²Laboratory for Robotic Research, Iran University of Science and Technology, Tehran, Iran. ³Students' Scientific Research Center (SSRC), Tehran University of Medical Sciences, Tehran, Iran. ⁴Nervous System Stem Cells Research Center, Semnan University of Medical Sciences, Semnan, Iran. ⁵School of Medicine, Iran University of Medical Sciences, Tehran, Iran. ⁶Guilan University of Medical Science, Guilan, Iran. ⁷Student Research Committee, School of Paramedical and Rehabilitation Sciences, Mashhad University of Medical Sciences, Mashhad, Iran. ⁸Student Research Committee, Zanjan University of Medical Sciences, Zanjan, Iran. ⁹Tehran Medical Sciences, Islamic Azad University of Medical Sciences, Tehran, Iran. ¹⁰Faculty of Medical Sciences, Canjan, Iran. ¹¹Shahid Beheshti University of Medical Sciences, Tehran, Iran. ¹²School of Medical Sciences, Cellege of Medicial Science, Zanjan, Iran. ¹³Student Research Committee, Iran University of Medical Sciences, Tehran, Iran. ¹⁴College of Medicine, University of Arizona, 1501 N Campbell Ave, Tucson, AZ 85724, USA. ¹⁵Fasa University of Medical Science, Fars, Iran.

Received: 28 February 2025 Accepted: 9 May 2025 Published online: 22 May 2025

References

- Moujalled D, Strasser A, Liddell JR. Molecular mechanisms of cell death in neurological diseases. Cell Death Differ. 2021;28(7):2029–44.
- Wheeler AC, Sacco P, Cabo R. Unmet clinical needs and burden in Angelman syndrome: a review of the literature. Orphanet J Rare Dis. 2017;12:1–17.
- Knopman DS, Amieva H, Petersen RC, Chételat G, Holtzman DM, Hyman BT, et al. Alzheimer disease. Nat Rev Dis Primers. 2021;7(1):33.
- Steinmetz JD, Seeher KM, Schiess N, Nichols E, Cao B, Servili C, et al. Global, regional, and national burden of disorders affecting the nervous system, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet Neurol. 2024;23(4):344–81.
- Huang Y, Li Y, Pan H, Han L. Global, regional, and national burden of neurological disorders in 204 countries and territories worldwide. J Glob Health. 2023. https://doi.org/10.7189/jogh.13.04160.
- Feigin VL, Vos T, Alahdab F, Amit AML, Bärnighausen TW, Beghi E, et al. Burden of neurological disorders across the US from 1990–2017: a global burden of disease study. JAMA Neurol. 2021;78(2):165–76.
- Deenen JC, Horlings CG, Verschuuren JJ, Verbeek AL, van Engelen BG. The epidemiology of neuromuscular disorders: a comprehensive overview of the literature. J Neuromusc Dis. 2015;2(1):73–85.
- Marneweck M, Hammond G. Voluntary control of facial musculature in Parkinson's disease. J Neurol Sci. 2014;347(1–2):332–6.
- Oliveira GC, Ngo QC, Passos LA, Oliveira LS, Papa JP, Kumar D. Facial expressions to identify post-stroke: a pilot study. Comput Methods Progr Biomed. 2024;250: 108195.
- Trevisan DA, Hoskyn M, Birmingham E. Facial expression production in autism: a meta-analysis. Autism Res. 2018;11(12):1586–601.
- 11. Davodabadi A, Daneshian B, Saati S, Razavyan S. Mathematical model and artificial intelligence for diagnosis of Alzheimer's disease. Eur Phys J Plus. 2023;138(5):474.
- Yolcu G, Oztel I, Kazan S, Oz C, Palaniappan K, Lever TE, et al. Facial expression recognition for monitoring neurological disorders based on convolutional neural network. Multimed Tools Appl. 2019;78(22):31581–603.
- 13. Jeong YU, Yoo S, Kim YH, Shim WH. De-identification of facial features in magnetic resonance images: software development using deep learning technology. J Med Internet Res. 2020;22(12): e22739.
- Jiang R, Chazot P, Pavese N, Crookes D, Bouridane A, Celebi ME. Private facial prediagnosis as an edge service for Parkinson's DBS treatment valuation. IEEE J Biomed Health Inform. 2022;26(6):2703–13.
- 15. Hughes RA, Brainin M, Gilhus NE. European handbook of neurological management. Amsterdam: Wiley; 2008.
- McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7(3):263–9.
- Roca F, Lang P-O, Chassagne P. Chronic neurological disorders and related comorbidities: role of age-associated physiological changes. Handb Clin Neurol. 2019;167:105–22.
- Ward C, Phillips M, Smith A, Moran M. Multidisciplinary approaches in progressive neurological disease: can we do better? J Neurol Neurosurg Psychiatry. 2003;74(Suppl 4):iv8–12.
- Volonté C. Machine learning as a new tool in neurological disease prevention, diagnosis, and treatment. Explor Neuroprotect Thera. 2023;3(1):1–7.

- Yolcu G, Oztel I, Kazan S, Oz C, Palaniappan K, Lever TE, et al., editors. Deep learning-based facial expression recognition for monitoring neurological disorders. In: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2017: IEEE.
- 21. Vieira S, Pinaya WH, Mechelli A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications. Neurosci Biobehav Rev. 2017;74:58–75.
- 22. Gautam R, Sharma M. Prevalence and diagnosis of neurological disorders using different deep learning techniques: a meta-analysis. J Med Syst. 2020;44(2):49.
- 23. Valliani AA-A, Ranti D, Oermann EK. Deep learning and neurology: a systematic review. Neurol Therapy. 2019;8(2):351–65.
- 24. Barbero-Gómez J, Gutiérrez P-A, Vargas V-M, Vallejo-Casas J-A, Hervás-Martínez C. An ordinal CNN approach for the assessment of neurological damage in Parkinson's disease patients. Expert Syst Appl. 2021;182: 115271.
- Berke Erdaş Ç, Sümer E, Kibaroğlu S. CNN-based severity prediction of neurodegenerative diseases using gait data. Digital Health. 2022;8:20552076221075148.
- Wang Y, Zhu J, Zhao J, Li W, Zhang X, Meng X, et al. Deep learning-enabled clinically applicable CT planbox for stroke with high accuracy and repeatability. Front Neurol. 2022;13: 755492.
- Huang J, Lin L, Yu F, He X, Song W, Lin J, et al. Parkinson's severity diagnosis explainable model based on 3D multihead attention residual network. Comput Biol Med. 2024;170: 107959.
- Lim WS, Chiu S-I, Wu M-C, Tsai S-F, Wang P-H, Lin K-P, et al. An integrated biometric voice and facial features for early detection of Parkinson's disease. NPJ Parkinson's Dis. 2022;8(1):145.
- Oliveira GC, Ngo QC, Passos LA, Papa JP, Jodas DS, Kumar D. Tabular data augmentation for video-based detection of hypomimia in Parkinson's disease. Comput Methods Progr Biomed. 2023;240: 107713.
- 30. Fan Y, Wang H, Zhu X, Cao X, Yi C, Chen Y, et al. FER-PCVT: facial expression recognition with patch-convolutional vision transformer for stroke patients. Brain Sci. 2022;12(12):1626.
- Bandini A, Rezaei S, Guarín DL, Kulkarni M, Lim D, Boulos MI, et al. A new dataset for facial motion analysis in individuals with neurological disorders. IEEE J Biomed Health Inform. 2020;25(4):1111–9.
- 32. Wang Y, Ye Y, Shi S, Mao K, Zheng H, Chen X, et al. Prediagnosis recognition of acute ischemic stroke by artificial intelligence from facial images. Aging Cell. 2024;23: e14196.
- Zhuang Y, McDonald M, Uribe O, Yin X, Parikh D, Southerland AM, et al. Facial weakness analysis and quantification of static images. IEEE J Biomed Health Inform. 2020;24(8):2260–7.
- Tripathi SC, Garg R. Consistent movement of viewers' facial keypoints while watching emotionally evocative videos. PLoS ONE. 2024;19(5): e0302705.
- 35. Zhu P, Wang H, Zhang L, Jiang X. Deep learning-based surface nerve electromyography data of E-health electroacupuncture in treatment of peripheral facial paralysis. Comput Math Methods Med. 2022;2022(1):8436741.
- Hou J-C, Thonnat M, Bartolomei F, McGonigal A. Automated video analysis of emotion and dystonia in epileptic seizures. Epilepsy Res. 2022;184: 106953.
- 37. Chen LY, Tsai TH, Ho A, Li CH, Ke LJ, Peng LN, et al. Predicting neuropsychiatric symptoms of persons with dementia in a day care center using a facial expression recognition system. Aging (Albany NY). 2022;14(3):1280–91.
- Singh G, Ramanathan M. Repurposing artificial intelligence tools for disease modeling: case study of face recognition deficits in neurodegenerative diseases. Clin Pharmacol Ther. 2023;114(4):862–73.
- Patel T, Othman AA, Sümer Ö, Hellman F, Krawitz P, André E, et al. Approximating facial expression effects on diagnostic accuracy via generative AI in medical genetics. Bioinformatics. 2024;40(Suppl 1):i110–8.
- 40. Karabegovic I. Artificial Intelligence in Industry 4.0: The future that comes true. 2024.
- Yousefi M, Akhbari M, Mohamadi Z, Karami S, Dasoomi H, Atabi A, et al. Machine learning based algorithms for virtual early detection and screening of neurodegenerative and neurocognitive disorders: a systematic-review. Front Neurol. 2024. https://doi.org/10.3389/fneur.2024.1413071.
- 42. Sharifi G, Hajibeygi R, Zamani SAM, Easa AM, Bahrami A, Eshraghi R, et al. Diagnostic performance of neural network algorithms in skull fracture detection on CT scans: a systematic review and meta-analysis. Emerg Radiol. 2025;32(1):97–111.
- 43. Steurer B, Vanhaelen Q, Zhavoronkov A. Multimodal transformers and their applications in drug target discovery for aging and age-related diseases. J Gerontol A Biol Sci Med Sci. 2024. https://doi.org/10.1093/gerona/glae006.
- Liu L, Liu S, Zhang L, To XV, Nasrallah F, Chandra SS. Cascaded multi-modal mixing transformers for Alzheimer's disease classification with incomplete data. Neuroimage. 2023;277: 120267.
- 45. Gkikas S, Tachos NS, Andreadis S, Pezoulas VC, Zaridis D, Gkois G, et al. Multimodal automatic assessment of acute pain through facial videos and heart rate signals utilizing transformer-based architectures. Front Pain Res (Lausanne). 2024;5:1372814.
- 46. Kopalidis T, Solachidis V, Vretos N, Daras P. Advances in facial expression recognition: a survey of methods, benchmarks, models, and datasets. Information. 2024;15(3):135.
- 47. Al-Hammadi M, Fleyeh H, Åberg AC, Halvorsen K, Thomas I. Machine learning approaches for dementia detection through speech and gait analysis: a systematic literature review. J Alzheimers Dis. 2024;100(1):1–27.
- Valliani AA, Ranti D, Oermann EK. Deep learning and neurology: a systematic review. Neurol Ther. 2019;8(2):351–65.
 Gautam R, Sharma M. Prevalence and diagnosis of neurological disorders using different deep learning techniques:
- a meta-analysis. J Med Syst. 2020;44(2):49.
 50. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021;8(1):53.
- Sajjad M, Ullah FUM, Ullah M, Christodoulou G, Alaya Cheikh F, Hijji M, et al. A comprehensive survey on deep facial expression recognition: challenges, applications, and future guidelines. Alex Eng J. 2023;68:817–40.
- 52. Ganesh S, Chithambaram T, Krishnan NR, Vincent DR, Kaliappan J, Srinivasan K. Exploring Huntington's disease diagnosis via artificial intelligence models: a comprehensive review. Diagnostics (Basel). 2023;13(23):3592.
- Iqbal S, Qureshi AN, Li J, Mahmood T. On the analyses of medical images using traditional machine learning techniques and convolutional neural networks. Arch Comput Methods Eng. 2023;30(5):3173–233.

- 54. Alfalahi H, Dias SB, Khandoker AH, Chaudhuri KR, Hadjileontiadis LJ. A scoping review of neurodegenerative manifestations in explainable digital phenotyping. NPJ Parkinson's Dis. 2023;9(1):49.
- Younis EMG, Mohsen S, Houssein EH, Ibrahim OAS. Machine learning for human emotion recognition: a comprehensive review. Neural Comput Appl. 2024;36(16):8901–47.
- Jafari M, Tao X, Barua P, Tan R-S, Acharya UR. Application of transfer learning for biomedical signals: a comprehensive review of the last decade (2014–2024). Inf Fusion. 2025;118: 102982.
- 57. Zhang J, Xie X, Peng G, Liu L, Yang H, Guo R, et al. A real-time and privacy-preserving facial expression recognition system using an Al-powered microcontroller. Electronics. 2024;13(14):2791.
- 58. Li Z, Li H, Meng L. Model compression for deep neural networks: a survey. Computers. 2023;12(3):60.
- Surianarayanan C, Lawrence JJ, Chelliah PR, Prakash E, Hewage C. A survey on optimization techniques for edge artificial intelligence (Al). Sensors. 2023;23(3):1279.
 Biska N, Use and Li W, Millettri F, Devisiti A, Bargarani G, et al. The feature of divisiti health with federated language.
- 60. Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. NPJ Digital Med. 2020;3(1):119.
- 61. Zhuang W, Chen C, Lyu L. When foundation model meets federated learning: motivations, challenges, and future directions. NPJ Digital Med. 2023. https://doi.org/10.1038/s41746-020-00323-1.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.