Social and Occupational Ergonomics, Vol. 152, 2024, 18–33
https://doi.org/10.54941/ahfe1005317

Human Factors and Facial Recognition Technology in Emergency Response: An Integrative Review

Aimee Kendall Roundtree

Texas State University, San Marcos, TX 78666, USA

ABSTRACT

Facial Recognition Technology (FRT) has the potential to enhance emergency response by improving efficiency, reducing response times, and potentially saving lives. It is important to explore both the benefits and challenges of implementing FRT in emergency response protocols and identify the gaps in research on its application in emergency response. This integrative review collected articles from 2010 onwards through keyword searches in titles and abstracts across multiple databases, with inclusion and exclusion criteria applied, focusing on English-language peer-reviewed articles mentioning specific keywords related to emergency services and facial recognition. Data collection involved converting PDFs to plain text and coding--both human and non-human--for thematic analysis refined over rounds of close readings for accuracy. The results reveal positive and mixed findings regarding FRT in emergency response across various studies. Several methodologies were employed, including machine learning and deep learning techniques, achieving high accuracy rates in identifying individuals, particularly in scenarios like disaster rescue and masked-face recognition during the COVID-19 pandemic. Studies also evaluated FRT's effectiveness in disaster victim identification, safety systems integration, and medical applications, showcasing its potential across different contexts. However, limitations such as challenges in real-world deployment, concerns regarding privacy and bias, and the need for further validation and standardization were highlighted across the studies, indicating areas for future research and development to enhance the technology's efficacy and ethical use. The review emphasizes the importance of addressing technical, ethical, and governance challenges to deploy FRT effectively and responsibly in emergency response, serving as a valuable resource for stakeholders and researchers seeking to understand and advance the field.

Keywords: Facial recognition technology, User experience, Ethics, Emergency response, EMT, Fire, Paramedics, Rescue, Disaster, Human factors

INTRODUCTION

Integrating artificial intelligence (AI) into emergency response might aid the capabilities of human responders and improve overall efficiency in crisis management. AI is transforming the way we approach disaster mitigation, enhance preparedness, reduce response time, and allocate resources. According to a report by the World Economic Forum (WEF) titled "Harnessing Artificial Intelligence for the Earth," the adoption of AI in emergency response may lead to a reduction in response times, thereby potentially saving countless lives and mitigating the impact of disasters on communities worldwide (WEF, 2022). However, it is imperative to explore the benefits, and challenges before widespread implementation.

Facial Recognition Technology (FRT) is one such AI application that might reshape the landscape of emergency response protocols for paramedics, emergency medicine technicians, and firefighters. Many studies have investigated FRT use in law enforcement, but questions linger about the technology's implementation in other forms of emergency response. Literature has also called for more studies on diverse FRT applications, more variety in FRT training sets, and more attention to the FRT user experience (Roundtree, 2021a; Roundtree, 2021b; Roundtree, 2023). FRT's capacity to identify people and access crucial medical information might help expedite treatment and enhance outcomes in critical situations. For example, in rescue efforts, FRT might be used to ensure that individuals in a building have been accounted for (National Academies, 2024). Using FRT linked to medical records, responders might be able to more quickly access info about medical histories, allergies, and pre-existing conditions. FRT might enhance the capabilities of frontline responders and optimize resources used during emergencies.

FRT is emerging as a transformative tool in emergency response, offering the potential for rapid and accurate identification of individuals in critical situations. This integrative review synthesizes the current research on human factors involved in facial recognition applications within the context of emergency response. The review examines FRT used in emergency-related environmental human factors and conditions to investigate its robustness and adaptability and help ensure reliable and human-centric performance in emergency scenarios.

METHODS

Articles from 2010 to the present were collected using keyword searches in titles and several mentions in abstracts. Scopus, ScienceDirect/Elsevier, Web of Science, PubMed, ERIC, IEEE Xplore, ACM, and Google Scholar. Eligibility criteria included inclusion and exclusion criteria for the review. Non-peer reviewed articles were excluded, as were patents and repeated listings. Only English-language articles were included. We also included articles that mentioned keywords (including facial recognition, fire, EMT, paramedics, emergency, and rescue) in the title and/or abstract but also that mentioned these terms in all subsections of the articles. Studies were grouped for the syntheses by theme. February 29, 2024, was the last date when each source was last searched or consulted. Scanning full texts was the method used to eliminate articles impertinent to the subject matter. The study met the inclusion criteria of the review when facial recognition and the requisite emergency services were mentioned in all subsections of the article. The review took place at three different time periods across three months to confirm the decision to include or exclude.

Methods used to collect data from reports included the conversion of PDFs to plain text documents, coding using NVivo, and categorization using

Orange topic modelling. NVivo is a qualitative data analysis software package designed to help researchers manage, analyze, and gain insights from qualitative data, commonly used in fields such as social sciences, psychology, market research, and others where qualitative data analysis is crucial. Orange is an open-source data visualization, machine learning, and data mining software package. It offers a user-friendly interface that allows users to interactively explore and analyze data without the need for extensive programming knowledge. Textual analysis included inductive coding and thematic categorization to identify themes pertaining to the science of engagement. Thematic analysis identifies and interprets patterns, themes, and meanings within an unstructured dataset. Labels or codes were assigned to segments of text that represent interesting features, concepts, or ideas. Segments were no smaller than one sentence and no larger than three or four sentences. Codes pertaining to issues and limitations were applied inductively moving from a specific instance to a general emerging theme. Codes were grouped together based on similarities and relationships to identify broader patterns or themes within the data. Themes were reviewed and refined over three close readings of the full text. Accuracy and limitations were tracked. Using both computer and human close reading helped minimize bias.

RESULTS

The studies reported overall positive or mixed results regarding FRT for emergency response. Of the 30 articles retrieved, seven were not peer-reviewed, three were repeats, three were not pertinent to the review, and 8 were patents. Nine studies were included in this review. Study methods included reviews (2), testing and training with datasets (2), user testing, protocol reporting, a case study, a pilot study, and a prospective observational study. See Table 1.

Al-Nabulsi et al. (2020) review image recognition techniques used during the COVID-19 pandemic as masked-face recognition systems for public health. It reviews machine learning and deep learning, for identifying individuals wearing masks and aiding in limiting the spread of the virus. Machine learning (ML) models, such as Support Vector Machines (SVM) and decision trees, are combined with deep learning (DL) models like Convolutional Neural Networks (CNN) to achieve high accuracy in identifying face masks. Deep learning techniques, particularly Convolutional Neural Networks (CNNs), MobileNets, and Naïve Bayes emerged as primary tools due to their ability to extract discriminative features efficiently. ML models were trained to classify images into categories such as “with mask” or “without mask” by collecting labelled images and training models to recognize patterns distinguishing between these categories. Techniques like CNNs, YOLO, and Residual Networks (ResNet) are applied for face mask identification. These models use features like ResNet 50 for extraction and multiple classifiers for improved accuracy. Specialized architectures like MobileNet v1 and MobileNet v2 are designed for embedded and mobile vision applications. Techniques for training and testing include supervised and unsupervised learning, cross-validation, and performance metrics such as accuracy, precision, recall, and F1 score.

In Al-Nabulsi et al. (2020), hybrid models combining multiple classifiers have demonstrated accuracy rates reaching up to 100% accuracy in some cases. Evaluations showed a range of accuracies, with CNN-based approaches outperforming others, while sensor-based networks also showed promising results. Furthermore, the availability of datasets, including the Real-World Masked Face Dataset (RMFD) and MaskedFace-Net, has facilitated the training and testing of the algorithms for improved performance across different scenarios and demographics. Hybrid deep transfer learning model incorporating SVM, decision trees, and combination techniques achieved 99.64% accuracy. Integration of YOLO-v2 and ResNet-50 DL (Residual Networks) achieved 81% accuracy. Lightweight neural networks for detecting people not wearing masks achieved 85% accuracy. A novel DL model using a public recognition database achieved 98% accuracy. Deep Learning and computer vision-based approaches achieved 95% accuracy. Multi-stage CNN architecture for face mask identification achieved 91.2% accuracy. CNN combined with a DL approach for mask identification achieved 95.8% accuracy. Multi-stage CNN architecture for face mask detection achieved 99.98% accuracy. Different types of deep learning for detecting face masks achieved 99.2% accuracy. CNN-based mask identification method utilizing OpenCV and MobileNetV2 achieved 99% accuracy. Deep learning using TensorFlow, Keras, and OpenCV achieved 99% accuracy. Lightweight Region Proposal Networks (RPNs) achieved 73% accuracy. Sensor Fusion (SF) approach achieved 99.26% accuracy. Smart Screening and Disinfection Walkthrough Gate (SSDWG) achieved 99.81% accuracy. Contactless sensors with computer vision achieved 91% accuracy. Sensors with deep learning achieved 97% accuracy.

In Broach et al. (2017), study subjects were selected from volunteers among faculty and staff members in a Department of Emergency Medicine and from students participating in health service career demonstrations. Injury patterns were categorized as mild, moderate, and severe based on the extent and combination of injuries. Victim moulage template sets were created with unique injury patterns. Photos were taken with various devices to simulate real-world scenarios. Pre-study photos were compared against moulage photos using facial recognition software, and statistical analysis was conducted on a data set of 106 participants. The sample included images of mostly white young faces (White $n = 81$ or 76.4%. Ages 15–25, $n = 28$ or 26.4%. Ages 26–35, $n = 21$ or 19.8%. Ages 36–45, $n = 29$ or 27.4%. Ages 46–65, $n = 27$ or 25.5%. Ages ≥ 66, $n = 1$ or 0.94%).

Correct matches between pre-study/no-moulage photos and moulage photos ranged from 39% to 49% for mild/moderate injury and 31% to 49% for severe injury in Broach et al. (2017). No statistically significant differences were found between devices (iPhone, iPad, digital camera) or severity of injury. However, there was a statistically significant difference in matching severe moulage between mobile phone photos and digital tablet photos, with mobile phones performing better in this regard. The mean percentage of correct matches ranged from 39% to 45%, depending on the device used. Correct match percentages exceeded 90% for optimal-quality photos. The average number of incorrect returns ranged from 5 to 7 for pre-study photos and 16 to 22 for no-moulage photos. Gender significantly affected correct matching, with males more likely to be correctly identified but also having a larger number of incorrect returns. Correct match percentages ranged from 39% to 49% for mild/moderate injury and 31% to 49% for severe injury. The mean percentage of correct matches ranged from 39% to 45%. Average number of incorrect returns ranged from 5 to 7 for pre-study photos and 16 to 22 for no-moulage photos. The study evaluated the accuracy of FRT in matching pre-study and no-moulage photos with moulage photos depicting mild, moderate, and severe facial injuries. The correct match percentages ranged from 39% to 49% for mild/moderate injuries and 31% to 49% for severe injuries across different devices. The mean percentage of correct matches ranged from 39% to 45%, depending on the device used. Higher percentages of correct matches (exceeding 90%) were observed for photos of optimal quality, indicating the importance of photo clarity and other factors affecting image quality. The average number of incorrect returns ranged from 5 to 7 for pre-study photos and 16 to 22 for no-moulage photos.

Dilip et al. (2022) propose a home security and child safety system that integrates various technologies such as gas sensors, ultrasonic sensors, and facial recognition. FRT would allow the system to classify authorized and unauthorized individuals attempting to enter the home. An unauthorized person would trigger system notifications and restrict access by not opening locks. The study involved participants (undescribed by the report other than "older adults") discussing and selecting robots suitable for different activities related to aging, such as healthcare, recreation, and commercial work. They acquired questionnaire responses from caregivers and conducted cognitive orientation examinations with control and experimental groups. The robot engages in discussions with users to gather feedback and reports back to caregivers with the results across 85 activities, categorized into basic activities of daily living (BADL), instrumental activities of daily living (IADL), enhanced activities of daily living (EADL), and social activities (SA). The system accurately identified individuals and triggered appropriate responses based on their classification. Comrade robots demonstrated higher accuracy responses compared to service robots across different activity types. Comrade robots were designed to facilitate interactions between humans and robots, aiming to enhance the quality of life for elderly individuals. In BADL, comrade achieves a remarkable accuracy rate of 99%. Similarly, in IADL, comrade exhibits a high accuracy rate of 92%. For EADL, comrade achieved an accuracy rate of 77%. In SA, comrade achieves an accuracy rate of 64%.

Gooroochurn et al. (2010) offered a registration framework for use in emergency neurosurgical procedures to simplify procedures for non-specialist medical personnel. The framework involves using different camera set-ups to extract and localize craniofacial landmarks in 3D space, from two to five cameras, depending on the required landmarks and views. Methods for placement of cameras involve frontal and profile views, and using cursor lines aligned to specific landmarks on the patient's face to ensure proper placement.. Gabor filters were used for feature extraction, and a polynomial neural network (PNN) was used for classification. A polynomial neural network served as the classifier, with network tolerance assessed for illumination changes and noise. Simulation studies using CT data and experimental work on an artificial skull tested the accuracy of the registration framework. Photogrammetry results with an artificial skull provided error estimates for comparison with simulation results. Detection rates of PNN for craniofacial landmarks extraction were evaluated using various databases, showing high accuracy. PNN tolerance to illumination variations and noise showed satisfactory performance under various conditions.

Gooroochurn et al. (2010) results show comparable error estimates between simulation and experimental work. Results of tests of the robustness of the PNN classifier indicate satisfactory performance under varying conditions. The detection rates for the extraction of specific landmarks, such as outer and inner eye corners and ear tragus, were high, ranging from 92% to 99% correct detection rates. The methodology showed robustness to illumination changes and noise, with the trained neural network performing well under various lighting conditions and noise levels. Simulation studies using CT data and projected views demonstrated acceptable registration errors, with maximum RMS errors ranging from 0.58 mm to 1.04 mm. Photogrammetry results with an artificial skull provided error estimates within acceptable ranges for practical applications. Detection rates were high for the PNN classifier extracting eye corners and ear tragus landmarks from images.

Khoo et al. (2020) review facial recognition used in disaster response for disaster victim identification (DVI) processes. Traditionally, DVI relies on methods like fingerprint analysis, dental records, and DNA comparison, which may face limitations in certain scenarios. Facial recognition offers an additional method for identifying disaster victims, especially in situations where antemortem records are scarce or unavailable. Research suggests that facial recognition has been employed in mass disaster scenarios like the Thailand tsunami, albeit with varying success rates. The technology's ability to capture facial images within the initial post-disaster period, before decomposition occurs, presents an opportunity to match victims' faces with existing government databases, aiding in positive identification. Considering its potential benefits in expediting the identification process and upholding the dignity of the deceased, some propose integrating facial recognition as the fourth primary identifier in DVI procedures, alongside traditional methods. This paradigm shift could enhance humanitarian forensic efforts, providing closure to families and aligning with the principles of organizations like the International Committee of the Red Cross (ICRC). It cites a study where 32.2% of Thailand tsunami victims were identified visually, indicating potential for facial recognition. The proposal suggests using AI as a primary identifier in DVI processes, aligning with humanitarian principles.

O'Neill et al. (2024) review Face Verification Service (FVS) trials conducted during the megafires in Australia aimed to assist in disaster relief efforts by using FRT to verify the identities of individuals applying for aid who had lost their identification documents. Service Australia (SA) staff considered the trial highly successful, with acceptable accuracy rates. The technical infrastructure developed by SA and the National Bushfire Recovery Agency (NBRA) was deemed effective and efficient. There were discussions about integrating FRT into future welfare infrastructure, foreseeing benefits for emergency situations beyond bushfires. The trial operated under a national state of emergency, suspending elements of the Privacy Act to allow for information sharing between SA and organizations like the Red Cross. However, specifics about biometric data sharing were unclear. In another context, during the COVID-19 pandemic, FRT was used in home quarantine apps for travellers in South Australia, Western Australia, and the Northern Territory. These apps were met with mixed reviews, with concerns raised about privacy and functionality. In the context of disaster victim identification (DVI) scenarios, researchers in Australia are exploring the use of FRT to compare post-mortem images with antemortem images for identification purposes. This work is conducted in anticipation of mass disaster scenarios.

Ramos (2022) reviews the use of biometric technologies in response to the COVID-19 pandemic. Contact tracing apps and health monitoring tools incorporate facial recognition as a means to track and control the spread of the virus. Despite the potential benefits for public health, concerns linger regarding the infringement on individuals' privacy and data protection rights. The evaluation focuses on specific examples from different countries, including China, Russia, and France, showcasing different approaches and levels of privacy protection. For instance, China's "Health Code Apps" uses facial recognition and personal data collection to assign quarantine statuses. The application raises concerns about surveillance and data misuse. Similarly, Russia's "Social Monitoring" app tracks individuals' movements and activities, with broad permissions and limited transparency regarding data processing. In contrast, France implements facial recognition for statistical analysis of mask-wearing compliance without centralized data storage.

Sivaprasad et al. (2022) piloted an intelligent home system primarily focused on enhancing safety, particularly for children and differently-abled individuals, by integrating various sensors and using Raspberry Pi for data processing and control, incorporating a gas sensor (MQ6) for detecting gas leaks, an ultrasonic sensor for motion detection, and a mobile application for real-time status updates and user interaction. The system includes a facial recognition system using the Caffe model to identify authorized individuals, further enhancing security. Gas Sensor (MQ6) response Time was less than 10 seconds The range of detection was 100 to 10,000 ppm. The facial detection using Caffe model response Time was also less than 10 seconds. The system showed the date, time, gas sensor reading (in ppm), and the intimation status in the mobile app when gas leakage was detected. It displays the status of the ultrasonic sensor, camera, user command inference, and the display status of the lock system. While the team provides detailed descriptions of the system's components and functionalities, the article lacks empirical evidence or results from testing or experimentation to validate the system's performance.

Zhang et al. (2022) uses deep learning technology to predict anemia in emergency department patients. The study trained a deep learning system using videos of six patients. The team conducted a prospective observational study in the critical care area of a hospital including adult patients. They conducted five-fold cross-validation, splitting the dataset into training and validation sets, and they set different class weights to handle class imbalance.

Model performance was evaluated using accuracy, sensitivity, specificity, and AUC. Two senior emergency department doctors assessed validation videos for anemia. Doctors' accuracy, sensitivity, and specificity were compared with the prediction model. Videos captured patients' faces under stable indoor lighting using a pad camera. The proposed framework consisted of three modules: video-image, face detection, and anemia prediction network. The study used Face++ by Megvii Co., Ltd. for face recognition and detection. The study compared five convolutional neural networks for anemia prediction, selecting InceptionV3 as the model. The protocol initialized models with pre-trained ImageNet weights that were fine-tuned on the dataset. It used data augmentation and class weights to address overfitting and class imbalance. Facial recognition techniques analyzed facial features, particularly membrane pallor, associated with anemia. Patients were mostly older men (Age mean ± SD 66.11 ± 17.51 anemic patients vs. 59.00 ± 18.06 non-anemic patients. Male 64.97% of 217 anemic patients vs. 57% of 99 anemic patients).

Table 1. Findings.

Article	Methods	Participants	Results	Limitations
Al-Nabulsi et al. (2020)	Review	Insufficient reporting	Average accuracy: 94.2%	Mask variability. Lighting and environmental factors. Rapid deployment. Computational resources. Database limitations. Ethical and privacy concerns. User acceptance.
Broach et al. (2017)	Testing and training with dataset	White, young faces	Correct matches ranged from 39% to 49% (mild/moderate injury) and 31% to 49% (severe injury).	Incorrect results. Level of injury insensitivity. Ideal conditions. External validity. Sample size and quality. No human in the loop.
Dilip et al. (2022)	User testing	Insufficient reporting	Basic activities: 99% accuracy. Instrumental activities: 92%. Enhanced activities: 77%. Social activities: 64%.	Generalizability. Sample size and representation. Technology limitations. Limited scope of activities. External validity. Data availability.
Gooroochurn et al. (2010)	Testing and training with dataset	Insufficient reporting	92% to 99% correct detection rates. Acceptable registration errors range from 0.58 mm to 1.04 mm.	Low accuracy requirement. Manual landmark selection. Experimental setting. External validity. Automated landmark extraction. Feature localization. Generalizability.
Khoo et al. (2020)	Protocol	Insufficient reporting	None	Sample quality. Postmortem issues. Database validation. External validity. Ethical and legal concerns. Affordability. Cultural differences.
O'Neill (2024)	Case study	Insufficient reporting	ID trial = highly successful. The home app = mixed reviews with concerns about privacy and functionality.	Data quality. Privacy issues. External validity and independent evaluation. Postmortem issues.

(Continued)

Table 1. Continued

Article	Methods	Participants	Results	Limitations
Ramos (2022)	Review	Insufficient reporting	Insufficient reporting	Insufficient regulations and safeguards. Data misuse. Private collaborations. Need for governance. Privacy concerns.
Sivaprasad et al. (2022)	Pilot study	Insufficient reporting	< 10 seconds facial detection using Caffe model response time	Limited scope. Theoretical framework. Generalizability. Affordability.
Zhang et al. (2022)	Prospective observational study	Older men	82.37% to 92.59% accuracy across different tasks	Sample selection bias. Exclusion criteria. Limited generalizability. Equipment dependency. Deep learning model selection. Data augmentation. Sample size. Clinician assessment. Ethnicity and diversity. Incomplete reporting.

The Zhang et al. (2022) system demonstrated high accuracy and sensitivity across these tasks. Of 362 patients recruited and 316 videos used for analysis, anemia was diagnosed in 217 patients based on complete blood count results. At the image level, the accuracy for the prediction model ranged from 82.37% to 92.59% across different tasks. When comparing the prediction model with clinical assessment, the accuracy of the prediction model was significantly higher than that of the senior doctors, with scores of 55.23% and 51.46% for the doctors. Sensitivity ranged from 66.61% to 92.59% for the prediction model at the patient level. Specificity varied from 32.51% to 69.23% for the prediction model at the patient level. scores demonstrate the comparison between the performance of the prediction model and the evaluation conducted by two senior doctors. The prediction model generally outperformed the doctors in terms of accuracy, sensitivity, and specificity in detecting anemia. (Accuracy: Prediction Model: 82.37% Doctor 1: 55.23% Doctor 2: 51.46% Sensitivity: Prediction Model: 92.59% Doctor 1: 66.61% Doctor 2: 63.20% Specificity: Prediction Model: 69.23% Doctor 1: 37.50% Doctor 2: 32.51%).

Positive and mixed results for accuracy aside, there were also several limitations of the studies. Regarding Al-Nabulsi et al. (2020), face masks come in various shapes, sizes, colors, and designs, posing a challenge for recognition algorithms to accommodate the diverse range effectively. Variations in lighting conditions, such as shadows, reflections, or poor illumination, can affect the visibility of facial features and the overall performance of face mask recognition algorithms. The urgent need for face mask recognition required quick deployment of technology, which posed challenges in developing algorithms and adapting them to different scenarios and environments. Implementing real-time face mask recognition systems that process large amounts of data quickly can be computationally demanding, requiring high-speed processing and response times. Training accurate and unbiased face mask recognition models requires diverse and representative datasets. The availability of such datasets and potential biases present in the data can impact the performance and fairness of the algorithms. Capturing and processing personal biometric data raises concerns about privacy, consent, and potential misuse of the collected information. Ensuring robust data protection measures and addressing privacy concerns are crucial for the ethical use of the technology. Face mask recognition systems often require user cooperation, such as proper positioning of masks, removing obstructions, or following specific guidelines. Achieving widespread user acceptance and compliance can be challenging, impacting the overall effectiveness of the technology.

Regarding Broach et al. (2017), photos were taken under ideal lighting conditions, possibly artificially enhancing software performance. Other factors like facial cleanliness or presence of perspiration were not assessed, which could affect real-world performance. The study did not account for non-subjects submitting photos in real-world scenarios, potentially leading to false-positive identifications. Unexpected results: The study found more incorrect results when using idealized "no-moulage" photos taken on the day of the study compared to the pre-study photos. The reason behind this discrepancy remains unclear, whether it was due to the similar background in all study photos or some other factor. Severe facial injury patterns would limit the utility of the facial recognition software. The study involved a small number of participants, which could limit its statistical power and generalizability. Manual verification would still be necessary due to several incorrect results accompanying each correct one. Using idealized "well photos" led to correct results roughly 90% of the time, suggesting that improving the quality of submitted photos could enhance the technology's utility. Future investigations with larger sample sizes would be beneficial in determining the actual utility of the technology in real disaster situations. Future research could explore the prevalence of false positives in such situations. Future studies with larger sample sizes would provide more robust insights into the technology's utility.

Regarding Dilip et al. (2022), the study focuses on a specific robotic system, so the findings might not be generalizable. The number of participants involved was not specified, so it was unclear if the sample adequately represents the target population. The effectiveness of the cognitive assessment method relies heavily on the technology used, so limitations or inaccuracies in the technology could impact the validity of the assessment results. Caregiver questionnaire responses may have introduced bias or inaccuracies. The study only focuses on a subset of these activities for data collection and analysis. The controlled experimental setting and specific conditions might limit external validity. Real-world applications were not further explored. The study does not provide specific details on how the datasets can be accessed. Further research and development should focus on enhancing the accuracy and reliability of facial recognition, as well as expanding the system's capabilities to incorporate additional sensors and functionalities for more security and convenience.

Regarding Gooroochurn et al. (2010), there was a relatively low accuracy requirement (less than 5mm). Some neurosurgical procedures may require higher precision. In both the simulation study and experimental work, all landmarks were manually selected, which might introduce subjectivity and potential human error. The experimental validation was conducted on an artificial skull, which may not fully capture the complexities and variations encountered in real clinical settings. The extent and rigor of this validation process are not clearly defined and do not establish standardized procedures for user validation to ensure consistency and reliability. The effectiveness of the automated methods, particularly in handling variations in image quality and anatomical structures, was not fully evaluated. The study primarily focuses on a specific set of neurosurgical procedures and may not be directly applicable to other medical specialties or surgical interventions. Detection rate was better for single faces with simple backgrounds than for more complex images with clutter and multiple faces. The current method may require manual intervention or refinement in this aspect.

Regarding Khoo et al. (2020), clear, high-quality images of the deceased are necessary, which, in situ, might not be possible in the aftermath of a disaster. Trauma to the body, decomposition, or damage to personal belongings would impact accuracy. Significant physical changes due to injury or post-mortem effects would pose a challenge. Existing databases or reference images for comparison may not always be comprehensive or up-to-date, especially for international victims or those with limited records. The success of facial recognition depends on having a database of reference images to compare against. In the aftermath of a large-scale disaster, obtaining accurate ante-mortem facial images for comparison can be challenging. The protocol does not discuss how it will measure internal or external validity, or how human intervention will assist interpretation and decision-making. Facial recognition algorithms can be prone to bias and inaccuracies, particularly when dealing with diverse populations or non-standard facial features. Ethical, legal, and privacy concerns surround the use of facial recognition technology, including issues related to consent, data protection, and potential misuse of sensitive information. Privacy concerns, consent issues, and potential misuse of data raise questions about the appropriateness of deploying facial recognition technology in this context. Variations in technology, databases, and protocols across jurisdictions can hinder the seamless exchange of information and coordination of identification efforts. Implementing facial recognition technology requires significant financial investment in equipment, software, and training. In regions with limited resources or infrastructure, deploying and maintaining facial recognition systems for disaster victim identification may not be feasible. In a large-scale disaster involving victims from diverse backgrounds, there may be challenges related to interoperability and standardization of facial recognition systems. Different cultures have varying attitudes towards post-mortem imaging and identification methods. Some communities may find the use of facial recognition technology for identifying the deceased disrespectful or culturally insensitive.

Regarding O'Neill (2024), the FRT system's effectiveness depended on clear biometric images, which could be challenging to obtain in chaotic disaster scenarios. Privacy and data security, especially with the suspension of privacy regulations during the national emergency, were concerns. The COVID implementation faced criticism for privacy breaches and technical glitches, highlighting broader ethical and practical concerns surrounding facial recognition technology. Privacy infringement and glitches in home quarantine apps might hinder broader acceptance and implementation of facial recognition technology in disaster management. The use of facial recognition technology raises ethical concerns regarding privacy, consent, and potential biases. There was a lack of independent evaluation, raising questions about its reliability. Post-mortem facial recognition poses unique challenges due to factors like decomposition and lack of standardized images, complicating comparisons with ante-mortem datasets and requiring innovative solutions to ensure accuracy and effectiveness. Decomposition of bodies complicates facial recognition accuracy, posing challenges for post-mortem identification.

Regarding Ramos (2022), the absence of specific legislation governing the processing, storing, or discarding of the collected data from these digital systems was noted. This lack of regulation leaves room for potential misuse of data after the crisis is over. Safeguards to protect people's data in both the short-term and long-term were unclear. It was unclear what measures are being considered to ensure that personal data collected is only used for addressing the spread of COVID-19 and not for other purposes. Worries about the potential misuse of the data collected by these systems were noted. Governments may not restrict FRT use after the crisis, and collected data might be used for purposes beyond the crisis. Some countries have collaborated directly with private companies to develop these digital solutions without sufficient oversight from legislative institutions or public discussion. This lack of transparency raises questions about the legality, accountability, and safeguarding of these systems. The European Commission and the European Data Protection Board provide guidance, but there is still a need for a comprehensive governance framework to regulate the adoption of facial recognition technologies and other digital tools during emergencies. The use of facial recognition systems and other digital tools raises significant privacy concerns. Clear privacy policies, information on data controllers, and disclosure of purposes for the use of personal data must address these concerns.

Regarding Sivaprasad et al. (2022), the study focused on the technical aspects of the proposed system, such as hardware components, sensor specifications, and software algorithms. It may lack in-depth discussion on broader implications, societal impacts, or potential ethical considerations associated with implementing such a system. It also lacks a thorough discussion of the theoretical frameworks or conceptual models guiding the research. There was no mention of extensive validation or testing procedures to assess the accuracy, reliability, or effectiveness of the proposed system, which limits whether the system can perform in real-world scenarios. The study did not discuss the scalability of the proposed system or its applicability to different contexts or environments. It also did not discuss ethical and privacy concerns, particularly regarding data security, consent, and potential biases. The overall affordability or accessibility of the proposed system, particularly for individuals or communities with limited resources, was unclear. There was limited discussion of the user interface design or user experience aspects of the proposed application.

Regarding Zhang et al. (2022), the study collected videos from patients in the critical care area of the emergency department of a single hospital, which may not be representative of the general population or of patients with different severity levels of anemia. Patients with certain conditions affecting facial color were excluded from the study, which could introduce bias, as those conditions may impact the appearance of anemia differently. The study population consisted of patients from a single hospital, potentially limiting the generalizability of the findings to other settings or populations. The study relied on a specific device for video collection and analysis. The performance and accuracy of the analysis may vary with different equipment or settings, limiting the reproducibility of the results. Choosing a particular model (InceptionV3) per performance metrics excluded other factors such as interpretability or computational efficiency that could impact use. The effectiveness of the data augmentation overfitting reduction techniques in capturing the variability of facial appearances associated with anemia may be limited. The study included 316 patients, which may not be sufficient to capture sufficient variability in facial appearances associated with anemia, especially for mild versus severe cases. Subjective assessments of the doctors may introduce variability and bias, but not in ways that fully capture the clinical utility or accuracy of the model in real-world settings. The study participants were all Chinese, which may limit the generalizability of the findings. Some details, such as the specific criteria used for anemia diagnosis or the methodology for face correction, were not fully described.

CONCLUSION

Findings cover techniques and methodologies to achieve high-precision identification in real-time emergency scenarios. FRT applications integrated artificial intelligence, machine learning, and deep learning algorithms to enhance the efficiency of facial recognition systems for emergency scenarios such as search and rescue, head injury treatment, natural disasters, accidents, and security incidents. Case studies, reviews, and empirical studies investigate human factors and contexts that emerge using facial recognition to aid first responders in swift and accurate victim identification, evacuation management, and overall situational awareness. Emergency responders propose using FRT to detect injury during rescue. FRT might help measure the cognitive load on emergency responders to address cognitive strain and facilitate rapid decision-making in critical situations. FRT might help identify patients and avoid misidentification, but the technology identified facial injuries only a third of the time if photos were not optimal. Emergency scenarios often involve suboptimal lighting conditions that obscure facial features and expressions. Facial recognition technology was effective in identifying craniofacial landmarks. FRT achieves more sensitivity (positives) than specificity (negatives) in identifying anemic patients. Using the technology has been proposed to protect small children separated from their parents without thoroughly evaluating FRT's efficacy in identifying children (Roundtree, 2021a). FRT has also been planned for robots for the elderly—for using facial cues to determine the health and pain state of the old person and for managing routine duties like navigating in the outside world in emergencies. FRT might help bring closure to families and protect the dignity of the deceased in identifying victims of disasters. Still, only a third of the cases were successfully identified visually, suggesting a high error rate.

The studies also suggested various challenges for facial recognition technology (FRT) in different emergency contexts. FRT algorithms face complexities in accommodating diverse face mask variations and lighting conditions, emphasizing the need for robust data sets and ethical considerations in deploying such technology. Issues with idealized photo conditions and the necessity for manual verification due to potential false positives limit external validity. Limitations of specific system's generalizability raised questions, as did the importance of improving accuracy and addressing privacy concerns. Questions remain regarding the effectiveness in disaster victim identification, especially regarding image quality, post-mortem changes, and ethical considerations. Privacy and reliability concerns persist with FRT implementation, particularly in disaster scenarios. Governance issues and technical aspects of FRT are unresolved, and striking a balance between internal validity strategies and human in the loop to overcome limitations and biases is essential, emphasizing the need for comprehensive validation and diversity in data sets.

Ethical considerations and privacy concerns emerged in implementing facial recognition in emergency response, as they have with applications in other industries (Roundtree, 2022; Roundtree, 2021c). Privacy, data ownership, protecting personal information, respect of person. and informed consent to opt-out are among the lingering ethical issues that persist in human factors emerging in FRT used in emergency response. The literature scrutinizes potential biases in facial recognition algorithms, the implications of false positives or negatives, and the protection of an individual's privacy rights. Ethical principles included lawfulness, transparency, proportionality, time-, scope-, and purpose-limited, secure, non-discriminatory, and inclusive of communities engaged, with protections and safeguards against abuse. Furthermore, questions linger about externalities with material consequences for emergency response, such as robotics and artificial intelligence eventually eliminating emergency response jobs. FRT has also been proposed at residential sites to increase resident autonomy and reduce staff burden, but not without increasing pressure on human workers to be creative, versatile, and prepared to meet new challenges and competition from FRT. Overall, these studies underscore the importance of addressing technical, ethical, and governance challenges in deploying FRT effectively and responsibly.

Overall, this integrative review evaluates the potential human-factors-related benefits and risks of rapid identification in emergency response and the need to protect privacy and individual rights. It also synthesizes best practices and frameworks for governance. Finally, it serves as a valuable resource for developers, policymakers, emergency responders, technologists, and researchers seeking an understanding of the current landscape, challenges, and future directions of facial recognition technology in the context of emergency response. Future research should include more user experience methodologies and user testing, as well as collaborations with industry partners to design in situ protocols (Roundtree, 2021d). More user testing, better reporting of methods and strategies, more studies that investigate social, governmental, and economic implications, more diverse sample demographics, and more consideration of the user experience are essential.

ACKNOWLEDGMENT

The study was supported by funding from the NEC Foundation and the State Farm Foundation.

REFERENCES

Al-Nabulsi, J., Turab, N., Owida, H. A., Al-Naami, B., De Fazio, R., & Visconti, P. (2023). IoT Solutions and AI-Based Frameworks for Masked-Face and Face Recognition to Fight the COVID-19 Pandemic. Sensors, 23(16), 7193.

Bradwell, H. L., Edwards, K. J., Baines, R., Page, T., Chatterjee, A., & Jones, R. B. (2023). Facial recognition lock technology for social care settings: A qualitative evaluation of implementation of facial recognition locks at two residential care sites. Frontiers in Digital Health, 5, 1066327.

Broach, J., Yong, R., Manuell, M. E., & Nichols, C. (2017). Use of facial recognition software to identify disaster victims with facial injuries. Disaster Medicine and Public Health Preparedness, 11(5), 568–572.

Dilip, G., Guttula, R., Rajeyyagari, S., Pandey, R. R., Bora, A., R Kshirsagar, P., & Sundramurthy, V. P. (2022). Artificial intelligence-based smart comrade robot for elders healthcare with strait rescue system. Journal of Healthcare Engineering, 2022.

Gooroochurn, M., Kerr, D., Bouazza-Marouf, K., & Ovinis, M. (2011). Facial recognition techniques applied to the automated registration of patients in the emergency treatment of head injuries. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 225(2), 170–180.

Hand, M. (2022). Facial Recognition Technology Coming To Virginia Hospital Center. Patch. https://patch.com/virginia/arlington-va/facial-recognition-technology-coming-virginia-hospital-center

Khoo, L. S., & Mahmood, M. S. (2020). Application of facial recognition technology on identification of the dead during large scale disasters. Forensic Science International: Synergy, 2, 238.

O'Neill, C. (2024). Disaster, facial recognition technology, and the problem of the corpse. New Media & Society, 26(3), 1333–1348.

Ramos, L. F. M. (2020, September). Evaluating privacy during the COVID-19 public health emergency: the case of facial recognition technologies. In Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance (pp. 176–179).

Roundtree, A. (2023). User Expectations of Facial Recognition in Schools and Universities: Mixed Methods Analysis. In: Tareq Ahram and Christianne Falcão (eds) Usability and User Experience. AHFE (2023) International Conference. AHFE Open Access, vol 110. AHFE International, USA.

Roundtree, A. K. (2022). Facial Recognition Technology Codes of Ethics: Content Analysis and Review. In 2022 IEEE International Professional Communication Conference (ProComm) (pp. 211–220). IEEE.

Roundtree, A. K. (2021a). Testing Facial Recognition Software for Young Adults and Adolescents: An Integrative Review. In International Conference on Human-Computer Interaction (pp. 50–65). Cham: Springer International Publishing.

Roundtree, A. K. (2021b). Correction to: Testing Facial Recognition Software for Young Adults and Adolescents: An Integrative Review. In HCI for Cybersecurity, Privacy and Trust: Third International Conference, HCI-CPT 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings 23 (pp. C1–C1). Springer International Publishing.

Roundtree, A. K. (2021c). Ethics and facial recognition technology: An integrative review. In 2021 3rd World Symposium on Artificial Intelligence (WSAI) (pp. 10–19). IEEE.

Roundtree, A. (2021d). Facial Recognition UX: A Case Study of Industry-Academic Partnerships to Promote User-Centered Ethics in Facial Recognition. In Proceedings of the 39th ACM International Conference on Design of Communication (pp. 240–246).

Sivaprasad, R., Chithra, V., Deepashri, R., Jeyashri, R., Prathibanandhi, K., & Subramanyam, M. N. (2022, December). An Astute Rescue System for Enhanced Security using Facial Recognition. In 2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS) (pp. 1–5). IEEE.

Zhang, A., Lou, J., Pan, Z., Luo, J., Zhang, X., Zhang, H.,... & Chen, L. (2022). Prediction of anemia using facial images and deep learning technology in the emergency department. Frontiers in Public Health, 10, 3917.