Indian Journal of Computer Science and Technology
Special Issue, Volume 5, Issue 1 (January-April 2026), PP: 113-119.
https://www.doi.org/10.59256/indjcst.20260501C018
International Conference on Cyber and AI Security (ICCAIS-2026)
www.indjcst.com
ISSN No: 2583-5300

RCNN Emotion Vision AI Powered Expression Analysis

Supratika Padhi[1], Manish Pradhan[2], Jayaguru sahu[3], Soumya Ranjan Mishra[4], Sachikanta Dash[5], Prahallad Kumar Sahu[6]

[1,2,3,4,6] Department of CSA, GIET University Gunupur, Odisha, India.
[5]Madanapalle Institute of Technology and Science University, Andhra Pradesh, India.

To Cite this Article: *Supratika Padhi[1], Manish Pradhan[2], Jayaguru sahu[3], Soumya Ranjan Mishra[4], Sachikanta Dash[5], Prahallad Kumar Sahu[6], "RCNN Emotion Vision Ai Powered Expression Analysis", Indian Journal of Computer Science and Technology Volume 05, Issue 01 (January-April 2026), PP: 113-119.*

Abstract: *Facial expression analysis are one of the primary methods of identifying emotions which scans as very interesting in its combination with the elements of psychology and technology. Facial emotion recognition Face recognition has been given a significant boost with the application of deep learning algorithms with the capacity to identify common emotions. A lot of progress has been achieved in the automatic facial emotion recognition (FER) in recent years. The technology has found use in many industries to enhance human machine interaction particularly in human-centred computing and the nascent emotional artificial intelligence (EAI) field. The research objective of researchers is to improve the functionality of systems in the process of identifying and discerning human facial expression and behavior in diverse scenarios. The contribution of the RCNN to the area has been intense due to the ample formation of these networks which led to the generation of different architectures as part and parcel of the endeavor to encounter increasingly complicated problems. This paper investigates the present progress of automated emotion recognition AER technology which uses computational intelligence methods for its development. The research presents field development through its examination of contemporary deep learning models which improved emotion detection results across various data types and real-world situations. It provides the summary of the most recent progress of the RCNN architecture of FER in the past decade, illustrating the mechanisms of collaboration of deep learning based methods and special databases to give the strongest results.*

Key Words: *RCNN, ML, DL, EAI, Emotion Detection, Computer Vision.*

I. INTRODUCTION

Human-computer is an interaction of technology with which the interface between people and computers is through physical or computer equipment. FRS is a system through which automatically recognize people. With the proper and solid FRS required for biometric research, it drives such active digital world race [1]. Recently, rapid advancements in the fields of pattern recognition and artificial intelligence have spurred a great deal more research into technologies for man-machine interaction. Facial Emotion Recognition and detection (FERD) is one such upcoming area where huge strides are developing in automatic translation software, machine-to-man communication, and so forth. While the paper emphasizes surveying and analyzing multiple features for extracting faces from an imagery, database creation for emotional content of an image, and prediction/classification in the existence off learner algorithm, and many more [2]. In general terms, "classical drives for FER can be broadly categorized into two classes. i.e., Feature Extraction and Recognition of Emotions [3]. Also dealing with images preprocessing and maintenance procedures [4]. At this stage, backdrop and non-face areas are omitted, and the face detection describes the limits of the face area by cropping it out. Most prevalent method of identifying emotions uses NN and other ml methodologies for the classification Retrieved features. Facial emotion recognition on the basis of facial actions is a very complex problem of recognition of automatically facial emotions states. It becomes impossible, then, to say that same emotional state would be communicated in exactly the same manner by different individuals. E.g: Mood of a person, skin color, age, and environmental influence affect the face. FER is generally carried out in three steps as illustrated in the figure: (i) Face Detection, (ii) Feature Extraction, and (iii) Emotion Classification [5].

The facial emotion detection and internship has three primary steps: it comprises the process of preprocessing, extraction of characteristics, and successful recognition of emotions. For the preprocessing, a number of methods are applied such as face recognition, removal, key image extraction, face landmark recognition etc. Face detection is based on many algorithms like Multi-Task Cascade Convolutional Network (MTCNN), Viola-Jones detector, Light and Fast Face Detector, Tiny Face Detection, FaceNet, Caffe-based face detector, Haar features-based cascade detector and Single Stage Headless (SSH) detector. Libraries like OpenFace, OpenCV, Dlib are used for landmark extraction and face detection, leading to extraction of the important face areas in the video by reducing the background noise [6].

While DNN technology is offering high accuracy levels for face recognition and face verification applications using images, it is an area of active research which is showing flexibility in the face of challenging conditions e.g. varied illumination, occlusions, facial expressions. These techniques exhibit good performance when applied to a large number of good image data, even images that are captured without a controlled environment with different light, pose and expression. However when images are effected by severe lighting changes, noise or low resolution, a significant degradation of the recognition accuracy is observed. Under such bad conditions, video-based approaches can help give insight into the dynamics of facial movements and can make recognition performance any better [7]. There have been recent advancements in facial recognition using deep learning techniques that seem to have got successful results. A notable approach is the Faster R-CNN, which has region suggestion and was initially put forward for object detection. In addition, some of the deep learning-based face recognition methods employ a sliding window approach, which slides the image at various scales and positions to detect the facial regions efficiently. Single shot detector (SSD), initially created for object detection has also been shown good in face detection tasks.

Techniques for facial features can be classified into geometric-based approaches, which represent face points from a geometric perspective to create feature vectors, and appearance-based techniques, such as Gabor wavelets, which extract features from focused or wide face images. Accurate feature extraction from one face to another is a challenging task, which is crucial for effective classification and analysis. The choice of features in FER has a significant impact on performance, making feature extraction an important and carefully considered step.

Feature extraction often involves the Facial Action Coding System (FACS), which monitors how facial muscles contract and relax at different intensity levels. FACS has been refined over time to improve its accuracy and robustness in recognizing subtle facial movements. Early 2000s research in the FER literature indicates that CNNs outperform multilayer perceptrons (MLPs) and provide robust results when handling scale variations and changes in face space, and excel at handling previously unseen changes in facial pose. CNNs, a widespread type of DL architecture, are recognized to capture high-level abstraction by using a hierarchical structure with multiple nonlinear representations and transformations.

The choice of an efficient loss function and the choice of a network architecture are two important factors to consider when designing convolutional deep neural networks (CNNs) for feature extraction. CNN architectures are broadly categorized as backbone in the computer science networks or as multi-network systems. Following their impressive results in ImageNet competitions, models with a more standard CNN architecture, such as SENet, VGGNet, AlexNet, ResNet N GoogleNet are widely studied models for researchers. These networks, and their variants, have been used extensively for face recognition. Furthermore, multi- structure networks have been developed to facilitate multi- task learning in which some face recognition task is performed concurrently with other tasks for some other purpose.

The first one is the preprocessing step as in Fig. 1. where a human face will be fed in and the facial parts on that face will be identified. In the second step, informative features will be elicited as per the various components of the face and upon which the model shall have to be trained under the final step proceed to create labels to the Emotions through the use of training data since a classifier will use the training data to do this [8]. There are groups of facial movement called Action Units (AUs). The classification of emotions is done in sets of Action Units. This is one of the machine learning methods that can be used to recognize the emotions and feelings based on facial expressions. Although Size of Data Deep Learning may affect performance.

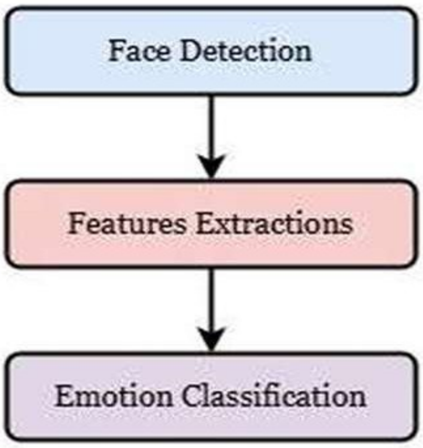

Fig. 1. Classification Step

II. LITERATURE REVIEW

It is very dutiful to indicate the opinions and moods of all mankind by facial expression. Indeed, there are many attempts to establish a system for conducting automatic facial expression analysis, since it is garnering wide applications in diverse fields (like in robotics, medicine, driving assist systems, and even in lie detection) [9].

Starting from the 20th century others had extensively studied so tremendously that they could categorize the 7 gesture (Glare, fear, smile, sadness, natural, disgust, and Amazement) and could define 7 primary gesture, regardless of the civilization where an individual grows. It speaks about mammoth research based on it on facial emotion recognition- related database components or the classifier for the facial emotion recognition and detection (FERD) research. Functioning on the visual features of images, some categorizer strategies are explained in [10].

There are two significant steps in faster R-CNN. The initial implementation phase known as Region proposal network (RPN) utilizing the new hype neural network Attention is suggested to enhance the Fast R-CNN. Faster R- CNN is much slower than Faster R-CNN and consumes a lot of time to generate candidates. The entire convolutional network operates as a complete system when it uses a Region Proposal Network (RPN) together with its components. The system provides full training capabilities which will generate candidate regions for the application to use thus delivering major time savings. The second stage of the process uses candidate regions which were identified by the earlier mentioned Region Proposal Network (RPN) to execute pooling operations and full linking functions before achieving final image recognition results [11].

That proves beneficial in UFQ that scrutinizes the methods of emotion recognition further. This paper studied future reactions based on reading images using a different classifier through emotion recognition. Basically, what are classification algorithms. Beyond that, several other issues like extreme makeup condition and expressions too have been attended to by the convolutional networks. Since then, computer vision and machine learning have greatly improved the success rate and ease of access to emotion recognition systems [12]. Hence, the research involvement in this field has been spurred. Here are a few applications involving this technology Human-Computer Interaction, psychiatric observation, identification of drunk drivers, and lie detection.

Recent improvements in face detection have yielded a success results using deep gaining knowledge of techniques . One remarkable technique is Faster R- CNN, which employs area proposals and turned into at the beginning delivered for object detection. Additionally, some deep learning-based face detection techniques appoint a sliding-window approach, scanning the photograph across various scales and positions to hit upon facial regions correctly. The Single Shot Detector (SSD), initially designed for object detection, has also verified effective in face detection applications [13].

Advances in facial emotion recognition (FER) were explored through the use of transfer learning (TL) in DCNNs. The authors introduced a novel FER method incorporating TL with a pipelined training strategy, resulting in an impressive accuracy of 89.52% on the JAFE dataset and 88.78% on the KDEF dataset, demonstrating the effectiveness of their approach. The architecture of CNN involves the use of pre-trained models with modified upper dense layers for emotion recognition, which effectively solves performance problems by exploiting learned features and reducing overfitting through fine-tuning [14].

CNN architecture was introduced in [15] to enhance recognition of facial expressions, particularly where there is a challenge in terms of occlusions and head tilt. The algorithm applied in this model was the Viola-Jones algorithm to detect faces, and local binary patterns (LBP) as the feature extraction algorithm. It showed remarkable accuracy rate of 82.66 percent and 84.94 percent in two experimental set-ups using the CK + and JAFFE data respectively. The architecture has been made up of five convolutional layers, a fully connected layer and softmax layer that are combined and can be used to efficiently extract features and categorize the emotions, thus enhancing the overall functioning of the system.

The findings in the works of C. Gautam et al., [16] indicate that the hybrid of handcrafted features and deep learning is effective and leads to the significant increase in the reliability of the emotion detection which is instrumental in a wide range of real-life applications.

The findings of X. Tao et al., [17] highlight the effectiveness of combining modalities for emotion detection, and provide a reliable framework that can be used in various real-time scenarios. Furthermore, the CNN architecture uses a multi-view fusion approach, which effectively captures complementary information from both modalities, addressing performance issues associated with single-modality systems. Emotion recognition has been improved by integrating insights of facial expressions with imaging photo plethysmography (IPPG) signals, in a multimodal framework. The research used machine learning methods which included SVM and K-Nearest Neighbor and Decision Tree and Random Forest to analyze IPPG signals and deep learning methods which included VGG16 and Vision Transformer to analyze facial expressions.

It investigated two strategies of mergers: decision level and functional level mergers. The preprocessing steps take in the video extracted from the face video and reconstruct the IPPG signal from the video and finally face detection steps using the Retina Face algorithm. The results show remarkable improvement in accuracy as the fusion on feature levels shows the accuracy of 72.37% and 70.82% for arousal and valence respectively, while SVM holds the highest accuracy for IPPG shown with 61.09% accuracy for arousal.

V. Manalu [18] focused on overcoming the challenge of distinguishing closely related emotions. The importance of these findings lies in their implications for improving human-computer interactions in various fields, including healthcare and consumer analytics.

III. METHODOLOGY

Affective Computation: in-depth Facial emotion recognition deals with sensing the various techniques and processes by which facial expressions convey emotions. In this regard, it may be taking the concept further by interfacing with computers, while human emotion recognition from video or audio data remains the area of concentration in most research activities [19]. Since this investigation appears to be restricted to face detection and recognition, it does not involve any convolutional neural networks for portraying emotion in the image. Emotion Recognition Introduction of their mechanism identifies the emotion plus the process that follows behind a certain technique and steps to comprehend the emotion.

Convolutional Neural Network (Fast-RCNN) was optimized by creating two mathematical models, each model is associated with a parameter which helps in debugging the network accuracy. The aim of the two models is to enhance the pace of the algorithm to its maximum, without compromising the Faster-RCNN algorithm, as well as execute the deep learning algorithm flawlessly on a mobile terminal that is not hardware-rich.

This is achieved successfully by eliminating a few data and a few parameters in the model. Primarily the fact that mathematical models, about which the Gaussian distribution is received, allow saving roughly 40 percent of the volume of the data is aimed at the fact that the algorithms of the convolutional neural networks can be successfully implemented in the compact robots, mini computers, and mobile devices like Raspberry Pi, and allow offer more quality to the human-computer interaction. Second, the last classification based on decision trees enables us to visualize the data more easily in order to trace the more weighted parameters.

Facial expressions, verbal channels, etc. were all used to judge emotions. Emotions play an important role in all walks of human life, and a lot of attention is being paid in worldwide research to emotion recognition [20]. Facial emotion recognition the psychomotor part of study aiming at recognizing human emotion from human facial expressions. Based on various surveys, the progress and advancements modern systems for emotion recognition equitably owe to such intricate algorithms. Recognizing the human emotions is complicated, as they may differ on the basis of wherein the emotion is being expressed, cognate cultures, or perhaps ambiguous facial giveaways [35-37]. Deep learning is machine learning in a large number of the way that the model is given learning to do on its own from data neighboring a specific task. Neural networks serve as the foundation for deep learning techniques, which find application in various fields including image recognition and image classification and decision making and pattern recognition [21]. The section presents a complete overview of the proposed method along with its emotional impact and all Inception model data used in the study.

The RCNN process needs RPN algorithm to determine which feature maps will be used as output results. The RPN algorithm identifies the target through its best fitting solution which defines the target identification box that most accurately covers the target area. The target recommendation process exists because the region suggestion networks support real-time object detection according to the demonstration in [22]. The algorithm positions the target optimally within the candidate box in terms of centrality.

The paper proposes the usage of classifier for human detection. It works simply with only one description feature vector consisting of linear, histogram, rectangle, and Euclidean features that classifier are trained with small features. In simple words the features represent grey-level changes corresponding to facial regions before a lot of skin areas were able to be skin-like, thus introducing some downsides [23]. One downside is slow computation for eigenvalues. To fasten its calculation, we adopted the method called integral graphs in this paper for-like features. Facial detection helps in the recognition of human expression [24]. After that, an image is prepared by decomposing it into the face and a non-face region.

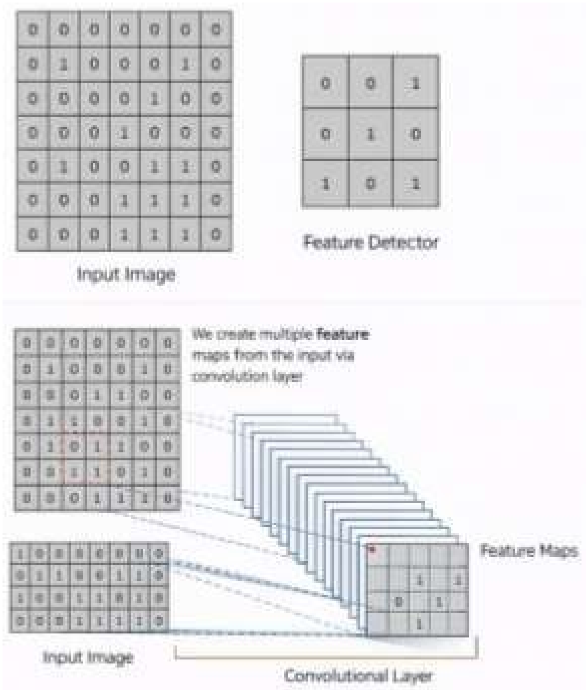

Fig. 2. Convolution Filter Operation

There are many possible ways to detect faces; however, a classifier basically calculates the feature by adding or subtracting across the dimensionality of the pixel group. The methodology itself should be applied and implemented by recognizing all the objects that contribute into the training stages of the classifier- a set of maximum elements that contribute to the problem of detection of the face-for the training stage of face detection, accuracy ranks very high with respect to computational complexity, it is on the contrary, low [25]. Feature Extraction is that process wherein the pixel data from the area of the face is converted to a higher-grade representation of its or its element features such as form, color, texture, and spatial configuration [26]. Dimensionality reduction of the input space such that it keeps the original essential information. Feature extraction plays an important role in developing an effective emotion classification as these facial features which are later fed into the classification module to classify emotions convey emotional states. Feature extraction will be [27].

Supervised Learning: A supervised learning type which instructs a system by use of labelled data. These data become a supervisor from which the model learns. Two types of supervised learning are classification and regression. [28] Neural Network (NN): NN maps by performing a non-linear mapping from the input into the lower dimension space. This means that it makes a statistical decision about the type of expression perceived, with each output unit predicting the probability of the corresponding input expression belonging to its particular class [29].

A standard database would need for any experiment in the domain of FER. One can say that the information is either primary or secondary. Just one primary dataset, in terms of demographics based on collecting the datasets, takes an eternity [30]. It is a well-known data sets for research in recognition of facial emotions. The data set has been used to evaluate many different deep learning architectures including convolutional neural networks (CNNs) among other architectural designs. The face dataset provides facial expressions for seven emotions which include happiness, sadness, anger, fear, surprise, disgust and neutral expressions[31-34].

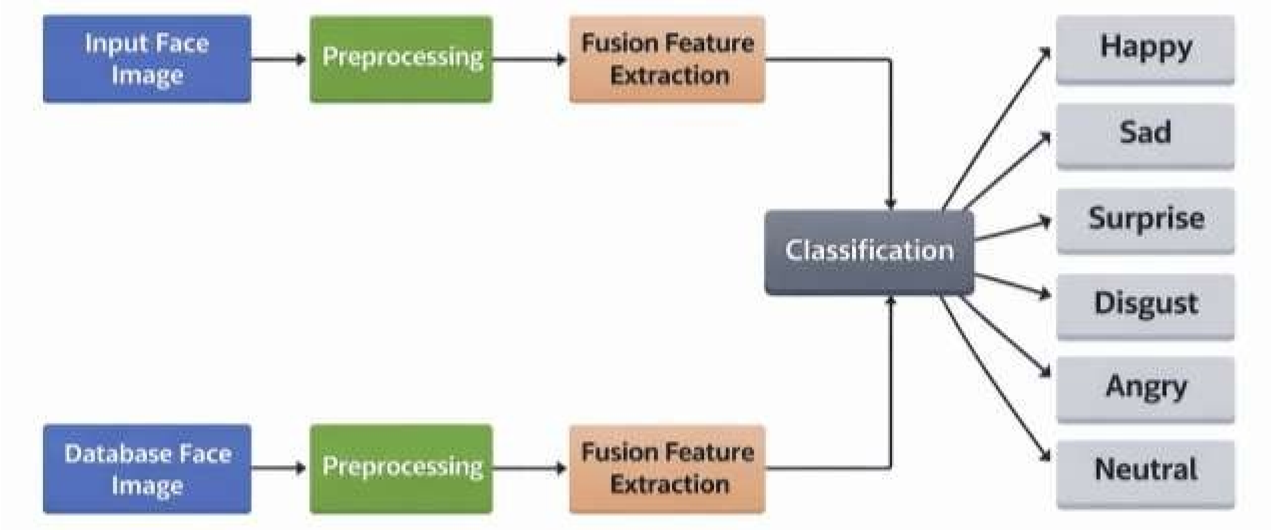

Fig. 3. Emotion Detection Process Model.

IV. RESULT ANALYSIS

The expression dataset was mainly used in performance testing the initial point of the algorithm. First, 414 sets of data were collected which included 79 sets of neutral expressions and 64 sets of sadness expressions and 72 sets of happiness expressions and 59 sets of disgust expressions and 78 sets of surprise expressions and 62 sets of anger expressions. The image data is as follows rise to the growing accuracy given that the more the images in the dataset, the higher the average accuracy (validation). The dataset used in this paper was split at 70:30 ratios, with 65% to 70% of the data used for training, while 25% to 30% was set aside for testing.

The system processes images through both Fast-RCNN and random forest optimization algorithm according to its two operational models. The system checks essential variables during its input stage to ensure that the data matches by verifying both the image and the computer system. The effects are divided into six categories: happy, sad, surprise, disgust, angry, neutral. Finally, two sets of facts are received primarily based on the two fashions.

Fig. 4. Facial Recognition Emotions Results.

The architecture of both RCNN for background subtraction (first-part CNN) and extraction of features from facial images and number of filters. Counting 1-8 layers in this paper, the highest accuracy was found for 4-layer count. Thus, direct correlation with respect to accuracy was with respect to the number of layers and inverse correlation with respect to execution time, which defied established norms.

The maximum accuracy was thus considering to 4-layers as it had achieved topmost accuracy value with respect to four-layer count. More layers increase the time for execution which we do not consider as an important factor in our study. Thus, the highest test set accuracy has been found through this technique over existing alternate methods. Most of the images hold peculiar poses, thus overall good reputation for the classification is maintained.

	Happy	Sad	Surprise	Disgust	Angry	Neutral
Happy	0.90	0.05	0.04	0.02	0.03	0.04
Sad	0.06	0.67	0.03	0.09	0.10	0.03
Surprise	0.03	0.03	0.86	0.06	0.03	0.05
Disgust	0.03	0.17	0.2	0.54	0.2	0.18
Angry	0.04	0.09	0.09	0.06	0.60	0.69
Neutral	0.06	0.09	0.03	0.06	0.05	0.80

Table I. Statistical Results

The experimental implications under condition of complete adherence to the conventional algorithm as well as the advanced algorithm. The categories of the experimental graphs, that is, the neutral, disappointment, happiness, worry, surprise, and anger, are indicated by the vertical axis of the desk, concurrently, the result form of the experimental version is being suggested by the horizontal axis, and the information is the percentage of a particular type of graph derived through the test belonging to exclusive varieties of results, i.e. the sum of the effects of the class. The diagonal facts are therefore of interest meaning that a version was accurate. To examine the difference between the units of information in a simpler way, a tool of evaluation bar chart is demonstrated below.

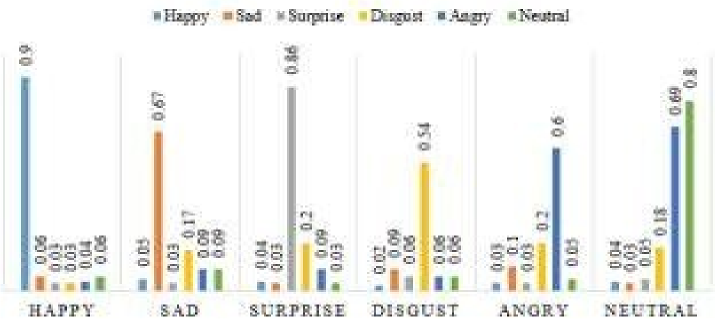

Fig. 5. Chart Bar of Statistical Results.

It is also protective from such variations and captures more in real life, whenever not frontal or angular capture scene. The other more complex situation was with multiple faces in one image equidistant from a camera, thus causing the algorithm not able to process it. Thus, more photons less accurate over fitting because less training photo lower accuracy. It also went further to say that needs tightly, for proper operation.

V. CONCLUSION

RCNN-based identification of facial expression, the most efficient way of obtaining facial features. The images of the training set are introduced directly into the network model as pixel value inputs. The extraction of the background greatly aids in the identification of emotions, the part of communication in which the real spirit of face-to-face human interaction lies. Such future developments will be tuned for useful applications to society at large and for human-robot interface in HRI. Increasing significance placed on emotion recognition will discuss the geometry of facial expression (eyes, eyebrows) against the laboratory image-based study on the implementation of real-life scenario tests and images taken from random scenes and get With new developments aimed at profile view trials, an entire range of applications will be opened, from patient monitoring in hospitals to security surveillance. Recognition of emotional expression may even go on to be combined with speech or body gesture for yet more applications to emerge in this new industry.

REFERENCES

[1] K. F. Azizan Illiana, "Facial Emotion Recognition: A Brief Review," in International Conference on Sustainable Engineering, Technology and Management 2018 (ICSETM-2018), 2020.
[2] Rath, L., Mishra, S. R., Dash, S., Pradhan, P. C., & Baboo, A. (2025). Predicting diabetic patients coronary artery calcium score, deep learning using retinal images. In Intelligent Computing Techniques and Applications (pp. 113-118). CRC Press.
[3] Pattnayak, A. P., Mishra, S. R., Dash, S., & Baboo, A. (2025). Utilization of deep learning and machine learning models to approach high glucose and low glucose prediction with type 1 diabetes mellitus in adult patients. In Intelligent Computing Techniques and Applications (pp. 102-107). CRC Press.
[4] Dora, N., Dash, S., Baboo, A., & Mishra, S. R. (2025, August). Efficient Nail Disease Diagnosis Using Deep Neural Networks for Predicting Abnormalities. In 2025 International Conference on Next Generation of Green Information and Emerging Technologies (GIET) (pp. 1-5). IEEE.
[5] Mishra, S. R., Dash, S., & Rath, L. (2024, November). Effective Diabetes Mellitus Prediction Using a Hybrid Ensemble Machine Learning Model with Iot. In 2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS) (pp. 1- 8). IEEE.
[6] Khuntuli, B., Dash, S., Pradhan, P. C., & Mishra, S. R. Combating food insecurity through remote sensing and machine learning for enhanced crop yield prediction. In Intelligent Computing Techniques and Applications (pp. 135-140). CRC Press.
[7] Sahu, P. K., Biswal, B. B., Mishra, S. R., Padhy, J., & Kumar, D. (2025, March). Demand-Based Secured Data Transmission in WSN. In

International Conference on Next Generation Computing and Communication Applications (pp. 37-44). Cham: Springer Nature Switzerland.

[8] R. Shyam, "Convolutional Neural Network and its Architectures.," Journal of Computer Technology & Applications, vol. 12, no. 2, pp. 6-14, 2021.

[9] Mishra, S. R., Dash, S., Padhy, S., Kumar, N., & Dash, Y. (2024, September). Integrating Multi-Omics Data for Advanced Diabetes Prediction and Understanding. In 2024 7th International Conference on Contemporary Computing and Informatics (IC3I) (Vol. 7, pp. 1447-1453). IEEE.

[10] Dash, S., Mishra, S. R., & Baboo, A. (2025, January). Enhancing Diabetes Prediction using Hybrid Ensemble Approach. In 2025 International Conference on Intelligent Systems and Computational Networks (ICISCN) (pp. 1-6). IEEE.

[11] R. Shyam, "Machine Learning and Its Dominant Paradigms," Journal of Advancements in Robotics, vol. 8, no. 2, pp. 1-10, 2021.

[12] Dash, S., Mishra, S. R., & Baboo, A. (2025, January). Efficient Prediction of Diabetes Mellitus Through Hybrid Ensemble Machine Learning Model Using IoT. In 2025 1st International Conference on AIML-Applications for Engineering & Technology (ICAET) (pp. 1-6). IEEE.

[13] Mishra, S. R., & Dash, S. (2024, December). Machine Learning Based Diabetes Prediction Using the PIMA Indian Dataset. In 2024 2nd International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES) (pp. 1-6). IEEE.

[14] M. A. H. Akhand, S. Roy, N. Siddique, M. A. S. Kamal, and T. Shimamura, “Facial emotion recognition using transfer learning in the deep CNN,” Electron., vol. 10, no. 9, 2021, doi: 10.3390/electronics10091036.

[15] A. S. Qazi, M. S. Farooq, F. Rustam, M. G. Villar, C. L. Rodríguez, and I. Ashraf, “Occlusions and tilt in facial recognition systems,” Appl. Sci., vol. 12, no. 1, 2022.

[16] C. Gautam and K. R. Seeja, “Facial emotion recognition using handcrafted features and CNN,” Procedia Comput. Sci., vol. 218, pp. 1295-1303, 2023, doi: 10.1016/j.procs.2023.01.108.

[17] X. Tao et al., “Facial video-based non-contact emotion recognition: A multi-view features expression and fusion method,” Biomed. Signal Process. Control, vol. 96, 2024.

[18] H. V. Manalu and A. P. Rifai, “Detection of human emotions through facial expressions using hybrid convolutional neural network-recurrent neural network algorithm,” Intell. Syst. Appl., vol. 21, p. 200339, 2024, doi: 10.1016/j.iswa.2024.200339

[19] H. Jiang and E. Learned-Miller, “Face detection with the faster R-CNN,” in Proc. 2017 IEEE Winter Conf. Appl. Comput. Vision,pp. 1-6, 2017.

[20] R. Shyam, "Automatic Face Recognition in Digital World," Advances in Computer Science and Information Technology (ACSIT), vol. 2, no. 1, pp. 64-70, 2015.

[21] Mishra, S. R., Dash, S., Padhy, S., & Das, R. K. (2024, September). Diabetic Foot Ulcer Diagnosis Through Deep Learning Model. In 2024 International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit) (pp. 1194-1199). IEEE.

[22] Mishra, S. R., & Dash, S. (2024). Iterative Model Design for Diabetes Analysis Using FedOmics Causal Network and Federated Multi-Omics Variational Autoencoder. International Journal of Emerging Technologies and Innovative Research, 11(6).

[23] S. R. N. S. M. A. H. Akhand, "Facial Emotion Recognition Using Transfer Learning in the Deep CNN," MDPI, vol. 10, no. 9, 2021.

[24] Mishra, S. R., & Dash, S. (2024). Predictive Analysis On Diabetes Detection Using Pima Indian Diabetes Dataset. IJRAR-International Journal of Research and Analytical Reviews (IJRAR), 11(2).

[25] Baboo, A., Mishra, S. R., & Dash, S. (2024, November). An Improvised Diabetes Prediction System Using Hybrid Ensemble Approach. In 2024 IEEE 11th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON) (pp. 1-6). IEEE.

[26] N. Mehendale, "Facial emotion recognition using convolutional neural networks (FERC)," SN Applied Sciences, vol. 2, no. 3, 2020.

[27] Mishra, S. R., & Dash, S. (2026). AI - Driven Remote Health Monitoring for Predicting Diabetes and Heart Diseases Using ULMCSO and PGND Models. Hyper - Intelligent Networks: Exploring the Future of Connectivity for Society 5.0, 219-247.

[28] Mishra, S. R., Dash, S., Padhy, S., & Samuel, P. (2026). Legal Aspects of Operating IoMT Applications in the Fog Computing. In Integrating Cloud, Fog, and Edge Computing in Healthcare: Federated Learning and Blockchain Approaches: Harnessing Distributed Technologies for Enhanced Healthcare Delivery (pp. 211- 224). Cham: Springer Nature Switzerland.

[29] Baboo, A., Dash, S., Padhy, S., & Samuel, P. (2026). Testing Perspectives of Fog-Based IoMT Applications with Federated Learning. In Integrating Cloud, Fog, and Edge Computing in Healthcare: Federated Learning and Blockchain Approaches: Harnessing Distributed Technologies for Enhanced Healthcare Delivery (pp. 199-210). Cham: Springer Nature Switzerland..

[30] N. R. S, "Emotion Recognition from Facial Expression using deep learning," International Journal of Engineering and Advanced Technology (IJEAT), vol. 8, no. 6S, 2019.

[31] Baboo, A., Patro, S. P., & Dash, S. (2024, December). A Deep Learning Approach for Enhancing Cardiovascular Disease Prediction Using ECG Data. In 2024 2nd International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES) (pp. 1-5). IEEE.

[32] Dash, A. B., Dash, S., Padhy, S., Kumar, N., Pati, G. K., & Uthansingh, K. (2025). Leveraging inception-v3 CNN model for efficient image classification. In Intelligent Computing and Communication Techniques (pp. 341-348). CRC Press.

[33] Dash, A. B., Dash, S., Padhy, S., Mishra, B., & Paikaray, B. K. (2025). Streamlining colorectal cancer diagnosis: leveraging MobileNet-V3 for efficient image classification. Int. J. Internet Manufacturing and Services, 11(4), 317.

[34] Dash, S., Padhy, S., Kumar, N., & Nayyar, A. (2026). Medisecure: a hybrid approach for enhancing multimedia data protection in healthcare. Cluster Computing, 29(1), 75.

[35] Dash, S., Padhy, S., Suman, P., Mal, S., Malviya, L., Suman, A., & Kishore, J. (2025). Privacy-preserving diabetes and heart disease prediction via federated learning and WCO. International Journal of Computational Intelligence Systems, 18(1), 217.

[36] Kumar, K. D., Dash, S., Ganiya, R. K., & Pradhan, P. C. (2025). A genetic algorithm-optimized hybrid CNN-LSTM for robust EEG seizure detection and adversarial defense. In Intelligent Computing Techniques and Applications (pp. 96-101). CRC Press.

[37] Khandakhani, S. W., Dash, S., Padhy, S., & Panda, R. (2024, December). Implementation of Customized Convolutional Neural Networks for Handwritten Marathi Character Recognition. In 2024 2nd International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES) (pp. 1-6). IEEE.