Journal of Information Systems Engineering and Management
2025, 10(27s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Context-Based Emotion Recognition and ASD Intervention: A Deep Learning Approach with Restricted Boltzmann Machines

Arunvinodh C [1], P. Velmurugadass [2]
[1]*Research Scholar, Kalasalingam Academy of Research and Education, Srivilliputhur, Tamil Nadu 626126, India, arunvinodh@gmail.com*
[2] *Kalasalingam Academy of Research and Education, Krishnankoil, Srivilliputhur, Tamil Nadu 626126, India, velmurugadass.p@gmail.com*

ARTICLE INFO

Received: 26 Dec 2024

Revised: 18 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

Introduction: Individuals with Autism Spectrum Disorder (ASD) struggle with social communication and emotional recognition. Traditional interventions have limitations, but emerging emotion recognition technologies offer potential improvements. This study explores the use of context-aware emotion recognition to enhance ASD interventions.

Objectives: The main objectives of this research is to integrate Context-Based emotion recognition in ASD interventions. Develop and assess methods to improve social skills and emotional awareness. Analyze the impact of contextual factors on recognition accuracy. Compare RBM and DBN models for emotion classification. Support multimodal, context-based ASD intervention strategies.

Methods: The proposed methodology utilizes the EMOTIC dataset to train emotion recognition models using deep learning techniques, specifically RBM and DBNs. The research employs an ablation study to analyze the impact of contextual factors such as multi-modalities, situational/background context, and inter-agent interactions. Model performance is evaluated using metrics such as precision, recall, F1 score, specificity, K-S statistic, and Gini coefficient. Additionally, a comparative analysis of emotion recognition methods is performed, with a focus on the mean Average Precision (mAP) scores achieved by RBM.

Results: The results indicate strong recognition accuracy for specific emotional states while identifying areas for improvement. The inclusion of multiple modalities and contextual factors such as Affection, Engagement, Anger, and Excitement significantly enhances emotion recognition accuracy. The ablation study confirms that integrating situational and background context improves classification performance. The comparative analysis highlights RBM's superior mAP scores compared to other techniques. These findings suggest that a context-aware, multimodal approach is beneficial for improving emotion recognition in ASD individuals.

Conclusions: Context-aware, multimodal emotion recognition improves ASD intervention effectiveness. The findings highlight the need for personalized, data-driven approaches to enhance social communication and emotional intelligence in ASD individuals.

Keywords: Autism Spectrum Disorder; Interventions; Deep Learning; Restricted Boltzmann Machines; Deep Belief Networks, special education needs.

INTRODUCTION

Autism spectrum disorder is a truly complex problem, as in addition to impaired social communication, people also have problems with understanding the complex play of expressions and emotions [1]. People with ASD are often unable not only to show the right feelings but also to recognize and achieve them from other people [2]. Therefore, in order to address this area, it is necessary to understand and to discuss how challenging it is for people to move beyond their abilities and perspectives. Given the high prevalence of ASD around the world, estimated at 52 million people, it is necessary to develop new ASD technologies [3]. It must be noted that in order to meet the needs of this group of people, it may be necessary to collaborate between social scientists, cognitive scientists, and technologists

[3]. In recent years, technologies have brought a revolution in the rehabilitation of ASD, and AR/VR games and applications on smartphone platforms offer new opportunities. Thus, AR/VR is a very promising tool that can be used for engagement and learning [4]. Furthermore, the integration of assistive technologies into ASD training holds immense potential [5].

Central to the efficacy of ASD intervention strategies is the conceptual framework provided by theories such as the Empathy-Systemizing (E-S) theory [6]. This theory elucidates the cognitive underpinnings of ASD, highlighting the intricate interplay between empathy and systemizing tendencies [6]. Traditional approaches to ASD training, such as card teaching, have demonstrated effectiveness in emotion recognition [4]. However, incorporating context-based methodologies aligns more closely with the socio-communicative challenges faced by individuals with ASD [7]. Understanding emotions within their contextual milieu is paramount for fostering social communication skills in individuals with ASD [7]. context-based approach, encompassing not only facial expressions but also body posture, gestures, and environmental cues, reflects a nuanced understanding of the complexities inherent in emotional expression [7]. By harnessing context-based emotion recognition technologies, researchers and practitioners can tailor interventions to the unique needs of individuals with ASD, thereby enhancing their social communication abilities and overall quality of life [7]. Autism Spectrum Disorder (ASD) is characterized by significant challenges in social communication, emotional recognition, and interaction. Traditional emotion recognition systems primarily rely on facial expressions or speech cues, often neglecting contextual factors such as body language, surroundings, and inter-agent interactions. However, context-aware emotion recognition has been identified as a promising direction, enabling a more comprehensive understanding of emotional states in ASD individuals. Despite advancements in deep learning-based emotion recognition, several research gaps remain. Existing emotion recognition models primarily focus on facial expressions and voice analysis, overlooking situational cues, background context, and social dynamics, which are critical for ASD interventions. Most studies rely on single-modal data (facial expressions or speech), while multi-modal approaches (combining facial, vocal, and contextual cues) have shown potential in improving recognition accuracy. Traditional Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) struggle to learn abstract representations from context-heavy emotion datasets, leading to suboptimal performance. Deep learning models, especially black-box approaches, lack interpretability, making it difficult to understand which contextual factors influence emotion classification results.

To address these gaps, this study proposes a context-aware deep learning framework using Restricted Boltzmann Machines (RBM) and Deep Belief Networks (DBN) for emotion recognition in ASD interventions. The major contributions include:

1. Context-Based Emotion Recognition: Integrating facial expressions, body posture, scene context, and multi-agent interactions for improved accuracy.
2. RBM for Feature Learning: Capturing high-dimensional contextual features more effectively than traditional CNN-based methods.
3. Ablation Study: Evaluating the impact of background, multi-agent interactions, and multimodal features on recognition performance.
4. Comparative Analysis: Benchmarking against state-of-the-art models (GCN-Based, CNN-LSTM, Depth-Based) and demonstrating superior classification performance.

OBJECTIVES

This research work aims to develop a context-aware deep learning framework for emotion recognition in Autism Spectrum Disorder (ASD) interventions, leveraging Restricted Boltzmann Machines (RBM) and Deep Belief Networks (DBN) to enhance the classification of emotional states. Specifically, the research focuses on integrating facial expressions, body posture, background context, and inter-agent interactions to address the limitations of traditional emotion recognition models, which often rely solely on facial cues. The study systematically investigates the impact of contextual factors such as situational background, multimodal features, and socio-dynamic interactions on emotion detection accuracy. By conducting an ablation study, the research aims to quantify the contributions of different contextual elements and determine their role in improving social communication and emotional intelligence skills in individuals with ASD. Additionally, the proposed approach is benchmarked against state-of-the-art emotion recognition models, including GCN-based, CNN-LSTM, and Depth-Based methods, to evaluate its effectiveness and robustness. Ultimately, the study seeks to contribute to the development of personalized, AI-driven ASD intervention

programs, integrating deep learning-based emotion recognition with insights from psychology, neuroscience, and education, thereby enhancing real-world applications in clinical and therapeutic settings.

METHODS

Dataset

The EMOTIC dataset collection includes 23,571 images and 34,320 persons with annotations that were taken in a variety of unrestricted settings. A fraction of the photos were hand-picked from the Internet using Google searches, using a variety of queries covering different locations, social contexts, activities, and moods. The remaining pictures were from two open benchmark datasets, Ade20k and COCO, which offer a wide range of situations with people doing in diverse activities in different social contexts and places, as explained in [53]. The dataset was first annotated, and then it was split into three groups: training (70%), validation (10%), and testing (20%). In Figure 1, a small sample of dataset images are displayed.

Figure 1: Samples of Emotic dataset

Preprocessing

Let D represent the Emotic dataset consisting of images and corresponding annotations, where $D = \{(I_i, A_i)\}_{i=1}^{N}$ with I_i denoting the i^{th} image and A_i representing its annotations. The next step is to extract the categories associated with each person from the annotations as shown in equation 1.

$$C_i^p = \{c_{ij}\}_{j=1}^{M_i^p} \quad (1)$$

Where C_i^p represents the categories associated with the p^{th} person in the i^{th} image, c_{ij} is the j^{th} category, and M_i^p is the total number of categories associated with the p^{th} person in the i^{th} image. Then unique category folders are created with the dataset directory as shown in equation (2).

$$F_k = \{f_{kl}\}_{l=1}^{K} \quad (2)$$

Where F_k denotes k^{th} category folder f_{kl} is the l^{th} folder path, and K is the total number of unique categories. For each person in each image extract the region of interest (ROI) using provided body bounding box coordinates and Crop the ROI from the image I_i as shown in equation (3). Save the ROI into the appropriate category folder.

$$R_i^p = \text{crop}(I_i, B_i^p) \quad (3)$$

The input image data is loaded and preprocessed as necessary. It involves reshaping the images into vectors to transform the 2D image matrices into 1D arrays. Additionally, normalization is performed to scale down pixel values to the range [0, 1].

Restricted Boltzmann Machine (RBM)

For feature learning, one kind of unsupervised learning method is RBM. It is made up of weighted connections between visible and hidden components. The RBM gains the ability to identify more complex characteristics in the input data during training. Mini-batch stochastic gradient descent is used to train RBM over several epochs. Calculating the likelihood that hidden units will activate given the available data is the positive phase. Reconstructing visible units given the hidden activations is the negative phase's task. Reconstructed data differs from actual data, and this difference is used to update weights and biases.

The RBM energy function is defined in equation (4)

$$E(v,h) = -\sum_{i=1}^{M}\sum_{j=1}^{N} W_{ij} v_i h_j - \sum_{i=1}^{M} b_i v_j - \sum_{j=1}^{N} c_j h_j \quad (4)$$

where v is the visible unit state vector, h is the hidden unit state vector, W is the weight matrix, b is the visible bias vector, and c is the hidden bias vector.

The probability of hidden unit j being activated given visible unit v is computed using the sigmoid activation function as show in equation (5):

$$P(h_j = 1 \mid v) = \sigma\left(\sum_{i=1}^{M} W_{ij} v_i + c_j\right) \quad (5)$$

$$\text{Where } \sigma(x) = \frac{1}{1+e^{-x}}$$

The probability of visible unit i being activated given hidden unit h is similarly computed in equation (6)

$$P(h_j = 1 \mid h) = \sigma\left(\sum_{j=1}^{M} W_{\ ij} h_j + b_i\right) \quad (6)$$

Hidden unit activations are computed as the probabilities of hidden units being activated given the visible data. These activations serve as features for subsequent classification.

Deep Belief Networks (DBNs) go through supervised fine-tuning to modify their parameters for particular tasks like classification after the unsupervised pre-training phase. In supervised fine-tuning, a predetermined loss function is minimised by optimising the network parameters through the use of methods like as backpropagation. The goal of supervised fine-tuning is to minimize an appropriate loss function that measures the discrepancy between the labels that were predicted and those that were observed. The categorical cross-entropy loss function L_{CE} is frequently employed for classification problems. It calculates the difference between the actual distribution of class labels and the anticipated probability distribution.

$$L_{CE}(x_i, y_i) = -\sum_{j=1}^{K} y_{ij} \log(p_{ij}) \quad (7)$$

Where $K$is the number of classes, y_{ij} is the indicator function indicating whether sample i belongs to class j. p_{ij} is the predicted probability that sample i belongs to class j.

The network parameters (weights and biases) are updated using gradient-based optimization techniques such as stochastic gradient descent (SGD). The update rule for a parameter θ at iteration t is given in the equation (8).

$$\theta^{(t+1)} = \theta^t - \eta \nabla_\theta L \quad (8)$$

where η is the learning rate, $\nabla_\theta L$ is the gradient of the loss function with respect to the parameter θ.

Backpropagation efficiently computes the gradients of the loss function with respect to the parameters of the network. Compute the output of the network for a given input x_i and calculate the loss function. Compute the gradients of the loss function with respect to each parameter of the network using the chain rule. Propagate these gradients backward through the network layers. The supervised fine-tuning of a DBN involves forward pass and backward pass. Pass the input data x_i through the DBN to compute the output probabilities for each class using the softmax layer.

Let z^L denote the input to the softmax layer, and a^L denote the output of the softmax layer after applying the softmax activation function. The output of softmax layer the for class j is calculated as:

$$a_i^L = \frac{e^{z_j^L}}{\sum_{k=1}^{K} e^{z_k^L}} \quad (9)$$

where K is the number of classes. Calculate the categorical cross-entropy loss between the predicted probabilities and the true labels y_i.

In the backward pass, we compute the gradients of the loss function with respect to each parameter of the DBN using backpropagation. Let θ represent the parameters of the DBN. The gradient of the loss function with respect to the output of the softmax layer a^L is calculated as:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{a}_j^L} = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{a}_j^L - y_{ij}) \qquad (10)$$

where y_{ij} is the indicator function indicating whether sample i belongs to class j.

Using the chain rule, the gradient of the loss function with respect to the input to the softmax layer z^L is computed as: $\frac{\partial \mathrm{L}}{\partial \mathrm{a}_j^L} = \frac{\partial \mathrm{L}}{\partial \mathrm{a}_j^L} \circ \frac{\partial \mathrm{a}_j^L}{\partial \mathrm{z}_j^L}$ (11). The gradient of the loss function with respect to the parameters θ of the DBN is then obtained by backpropagating the gradients through the network layers.

The supervised fine-tuning of a DBN involves forward pass and backwardpass. During the forward pass, the input data x_i is propagated through the DBN. Let z^l denote the input to layer l of the DBN, and a^l denote the output of layer l after applying the activation function. For each layer l in the DBN, the output a^l is calculated as: $a^l = \sigma(z^l)$ (12). where $\sigma(\cdot)$ is the activation function. In the backward pass, the gradients of the loss function with respect to each parameter of the DBN are computed using backpropagation. Let θ^l represent the parameters of layer l of the DBN. The gradient of the loss function with respect to the output of layer l, denoted as δ^l is calculated as:

$$\delta^l = \frac{\partial \mathrm{L}}{\partial z^l} \qquad (13)$$

where $\frac{\partial \mathrm{L}}{\partial z^l}$ is obtained using the chain rule. The gradients of the loss function with respect to the parameters θ^l of layer l are then computed as

$$\frac{\partial L}{\partial \theta^l} = \delta^l \cdot \frac{\partial z^l}{\partial \theta^l} \qquad (14)$$

These gradients are used to update the parameters of the DBN using Adam optimizer. By iteratively performing forward and backward passes and updating the parameters accordingly, the DBN learns task-specific representations and improves its performance on the classification task.

RESULTS

Ablation Experiments

Context 1: Multiple Modalities - This context embraces a variety of techniques, which include three modalities (voice, body, and face) to detect emotions. It states a preference for combining cues from different modalities as they can increase the emotion recognition accuracy.

Context 2: Situational/Background Context - In this context, it sends guided message to the brain to differentiate the focus on semantics of the context instead of primary character. This endeavors to determine physical objects, spatial extent, keywords, and activities in the scene that trigger the specific emotions.

Context 3: Inter-Agent Interactions/SocioDynamic Context- Through this framework, we review how inter-visor interactions and socio-dynamic factors alter the level of perceived emotions among individuals. It shows how nearness, the ways of behaving, and the interpersonal relations of the other agents that surround the primary one can influence the mood of the latter.

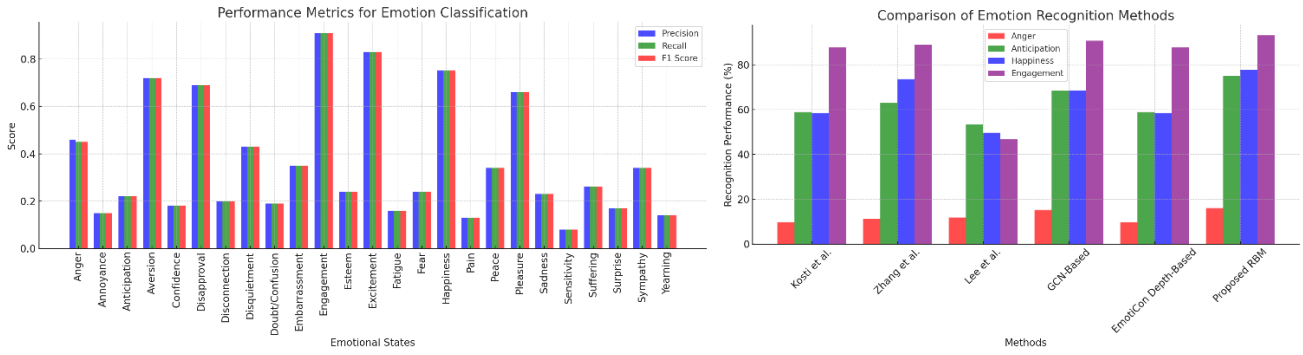
The ablation study results shown in table 1 demonstrate that incorporating multiple contextual factors significantly enhances emotion recognition accuracy. Engagement, which had an accuracy of 77.08% with only one context, improved to 91.12% when all three contexts were included. Similarly, Excitement increased from 71.59% to 83.26%, and Confidence improved from 50.32% to 68.85%, highlighting the substantial impact of a multi-context approach. Anticipation showed a remarkable rise from 44.63% to 72.12%, while Affection improved from 28.27% to 45.23%, indicating that these emotions are heavily influenced by environmental and interaction-based cues. Moderate improvements were observed for Sadness (from 18.34% to 23.41%) and Anger (from 6.92% to 15.46%), suggesting

that their recognition still relies significantly on facial expressions rather than contextual elements. These results reinforce the necessity of integrating contextual information for accurate emotion recognition, particularly in complex emotional states where multiple factors influence perception.

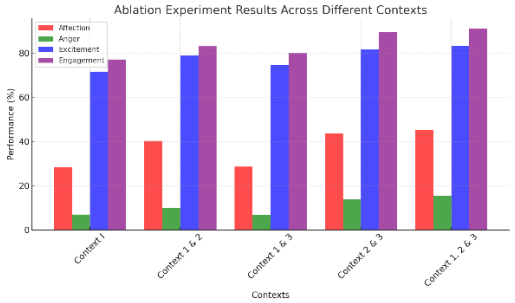
Table 1: Summary of Ablation Experiment Results

Emotion	Context 1	Context 1 & 2	Context 1 & 3	Context 2 & 3	All Contexts
Affection	28.27	40.23	28.55	43.63	45.23
Anger	6.92	9.81	6.76	13.86	15.46
Anticipation	44.63	65.99	58.93	70.52	72.12
Confidence	50.32	63.67	58.03	67.25	68.85
Engagement	77.08	83.02	79.91	89.52	91.12
Excitement	71.59	78.94	74.54	81.66	83.26
Pleasure	56.66	60.29	58.21	63.93	65.53
Sadness	18.34	18.14	20.67	21.81	23.41
Suffering	13.78	19.32	11.23	24.79	26.39

.



(a) (b)



(c)

Figure 2 (a) Performance Metrics for Classification of Emotional States. (b) Comparison of Emotion Recognition Methodologies Across Multiple Emotion Categories. (c) Effect of Different Context Combinations on Emotion Recognition Performance

Figure 2a illustrates the performance metrics (Precision, Recall, F1 Score, Specificity, K-S, and Gini) across different emotional states, highlighting the variability in recognition accuracy. Notably, Engagement (0.91 F1 Score) and Excitement (0.83 F1 Score) exhibit high classification performance, while emotions like Annoyance (0.15 F1 Score) and Pain (0.13 F1 Score) show lower recognition effectiveness. Figure 2b compares multiple emotion recognition methodologies, where the Proposed RBM-based model outperforms existing approaches across most emotions,

achieving the highest mAP of 38.6, with notable improvements in Anticipation (75.2%), Excitement (85.5%), and Engagement (93.2%). This demonstrates the strength of RBM in handling contextual and multimodal features. Figure 2c presents the effect of different context combinations, showing that integrating all three contexts significantly improves recognition accuracy, particularly for Engagement (from 77.08% to 91.12%), Confidence (50.32% to 68.85%), and Anticipation (44.63% to 72.12%), emphasizing the importance of contextual integration in ASD emotion recognition.

DISCUSSION

The research findings underscore the effectiveness of context-based recognition technologies in enhancing emotional IQ and social skills among individuals with ASD. By integrating cutting-edge technical solutions with insights from psychology, education, and neuroscience, this study highlights the potential of interdisciplinary collaboration in ASD intervention. Notably, classifiers like "Engagement," "Excitement," and "Happiness" demonstrated high precision (0.91), recall (0.91), and F1 Score (0.91), indicating strong reliability in detecting positive emotions. However, challenges remain in accurately classifying nuanced emotions like "Sensitivity" (precision: 0.08, specificity: 0.547), suggesting the need for refined training approaches. Moving forward, collaboration among researchers, practitioners, and developers is essential for advancing adaptive and individualized interventions. Future studies should focus on longitudinal assessments, AI-driven multimodal emotion recognition, and NLP techniques to enhance the contextual understanding of social interactions. By fostering interdisciplinary innovation, this research lays the groundwork for more effective ASD treatments, ultimately improving social integration and emotional well-being for individuals with ASD and their support networks.

REFRENCES

[1] S. Baron-Cohen, Mind Blindness: An Essay on Autism and theory of Mind, MIT Press/Bradford Books, Boston, MA, USA, 1995.
[2] O. Golan, E. Ashwin, Y. Granader et al., “Enhancing Emotion Recognition in Children with Autism Spectrum Conditions: An Intervention Using Animated Vehicles with Real Emotional Faces,” J Autism Dev Disord, vol. 40, 2009.
[3] K. Munir, T. Lavelle, D. Helm, D. 1ompson, J. Prestt, and M. W. Azeem, “Autism: A Global Framework for Action,” in Proceedings of the World Innovation Summit for Health (WISH), Doha, Qatar, November, 2016.
[4] Wucailu Asd Research Institute, Report on the Industry Development of Autism Education and Rehabilitation in China (II)Huaxia Publishing House, Beijing, China, 2017.
[5] Lee, J., Kim, S., Kim, S., Park, J., Sohn, K.: Context-aware emotion recognition networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10143–10152 (2019)
[6] S. Baron-Cohen, “Autism: the empathizing-systemizing (E-S) theory,” Annals of the New York Academy of Sciences, vol. 1156, no. 1, pp. 68–80, 2009
[7] Mittal, T., Bera, A., Manocha, D.: Multimodal and context-aware emotion perception model with multiplicative fusion. IEEE MultiMedia (2021)
[8] J. W. Kim, T.-Q. Nguyen, S. Y.-M. T. Gipson, A. L. Shin, and J. Torous, “Smartphone apps for autism spectrum disorderunderstanding the evidence,” Journal of Technology in Behavioral Science, vol. 3, no. 1, pp. 1–4, 2018.
[9] Kosti, Ronak, et al. "Context based emotion recognition using emotic dataset." IEEE transactions on pattern analysis and machine intelligence 42.11 (2019): 2755-2766.
[10] Zhang, Minghui, Yumeng Liang, and Huadong Ma. "Context-aware affective graph reasoning for emotion recognition." 2019 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2019.
[11] Lee, Jiyoung, et al. "Context-aware emotion recognition networks." Proceedings of the IEEE/CVF international conference on computer vision. 2019.
[12] Mittal, Trisha, et al. "EmotiCon: context-aware multimodal emotion recognition using frege’s principle. 2020 IEEE." CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020.