

---

## Chapter 8

# Conclusion and Future Scope

This chapter summarizes the main contributions of the thesis. The conclusion of the thesis is presented in Section 8.1, limitations are described in Section 8.2, and the future scope of the research work is presented in Section 8.3.

### 8.1 Summary and Conclusion

Recent advancements in technology have created numerous opportunities to provide learners with diverse devices and solutions that enhance the learning experience. Assessing learners' engagement in TEL through a mulsemmedia-enhanced approach is essential to maximize learning outcomes, improve learner satisfaction, and optimize educational resources and strategies tailored to learners' needs. However, existing mulsemmedia-based approaches have not significantly assessed learner engagement levels. This research proposes a FER approach to analyze learner engagement and satisfaction in mulsemmedia-based learning.

Current FER systems, however, face challenges such as overfitting and difficulties in accurately predicting expressions in real-time environments. Specific limitations of existing FER systems include the following: detecting faces under varying head poses, illumination conditions, and occlusions remains challenging, especially with the Viola-Jones face detection algorithm. Additionally, while most FER systems rely on spatial features, real-time FER recognition requires both spatial and temporal features, which current models often lack. Furthermore, supervised FER approaches require extensive datasets to train models effectively, and imbalances in facial expression counts can further impact model performance. To address these issues, enhanced FER techniques have been proposed in this thesis, which are summarized as follows.

- **Enhancing Viola-Jones Face Detection Prediction Accuracy:** The conventional Viola-Jones algorithm often struggles with detecting faces due to the exhaustive search required by AdaBoost, which evaluates all possible thresholds for Haar-like features to minimize training error. This process is time-consuming and can hinder the efficiency of feature selection and classifier optimization. To improve upon this, we

proposed enhancing the Viola-Jones algorithm by integrating PSO. Our approach leverages PSO in two key areas: first, to dynamically select optimal threshold values for feature selection, which enhances computational efficiency; and second, to refine the feature selection process by integrating PSO with AdaBoost to identify the most discriminative features for building a robust classifier. This enhancement significantly reduces the time and complexity associated with feature selection, particularly in challenging face-detection scenarios. The proposed approach achieved a significant average true positive rate of 98.73% and a 2.1% increase in average prediction accuracy compared to both the conventional Viola-Jones algorithm and current state-of-the-art methods.

- **FER Using CNN-BiLSTM Architecture:** Existing FER approaches are inadequate for analyzing spatiotemporal features in real-time environments involving dynamic facial movements. To address this limitation, we introduced a deep learning approach that combines CNN and Bi-LSTM networks. Our method uses a hyperparameter-tuned VGG-19 model structure with a time-distributed layer to automatically extract spatial features from sequences of images, these features are then processed by a Bi-LSTM to capture temporal features in both forward and backward directions, allowing for effective recognition of emotions from sequences of expressions. The proposed method experimented with a CK+ benchmark dataset and an In-house dataset. The approach demonstrated accuracies of 92% on the CK+ dataset and 84% on the In-house dataset, surpassing baseline methods and state-of-the-art techniques.
- **FER Implementation with Deep Semi-supervised Convolutional Sparse Autoencoder:** Deep Semi-supervised Convolutional Sparse Autoencoder has been introduced to address the limitations of supervised FER approaches, particularly their reliance on extensive datasets and the challenges posed by imbalanced facial expression counts, which can affect model performance. This approach involves two stages. In the first stage, a deep convolutional sparse autoencoder is trained with unlabeled facial expression samples, introducing sparsity in the convolutional layer to focus on relevant features. In the second stage, the encoder weights are fine-tuned with labeled samples using a fully connected layer and softmax activation for emotion

classification. This approach is experimented on the CK+, JAFFE, and In-house datasets, achieving high accuracy rates of 98.98%, 93.10%, and 98% respectively. Additionally, the performance and predictions of the proposed model were further analyzed using XAI techniques, such as Grad-CAM and image-LIME, to enhance interpretability.

- **Analysing Learner Engagement in Multimedia Synchronized Learning Environment:** We developed an IoT-based multimedia-synchronized learning platform using affordable components, including cooling fans, humidifiers, and haptic devices, to create an immersive multisensory learning environment. This system enhances the learning experience by incorporating sensory effects, such as the aroma of rosemary synchronized with rosemary-related content, and vibrotactile and airflow effects aligned with thunder and lightning learning content. These effects are seamlessly integrated with traditional audiovisual learning content. The results demonstrated that multimedia-based learning significantly improved learning outcomes and heightened enjoyment levels, providing learners with an enhanced sense of reality. Furthermore, we analyzed learners' engagement levels using the proposed FER approach within this context. The findings reveal that learners were more actively engaged in multimedia-based learning compared to conventional learning methods.

## 8.2 Limitations

Although the proposed FER techniques have shown promising results in real-time environments for analyzing learner engagement levels, certain limitations remain due to resource constraints. Some of these limitations are listed below:

- Focusing only on female facial expressions in data collection to train a deep learning model may lead to biased results when including other genders and may have limitations in real-world applications.
- Facial expressions alone may not accurately reveal learners' true emotions. Therefore, incorporating physiological signals could improve the accuracy of identifying learners' emotional states, enabling a more personalized learning experience.

- Cultural and age differences can influence the accuracy of facial expression prediction, as the proposed model was trained on FER datasets with samples from specific cultural backgrounds. Accuracy could be enhanced by incorporating more diverse FER samples, although careful mechanisms are needed when combining FER data from different cultures.
- Facial expression with high illumination, full occlusion, and extreme head pose variation led to lower emotion prediction accuracy in real-time environments. Accuracy in these challenging conditions can be improved through specialized transformers and self-attention techniques.

### 8.3 Future Scope

There is significant potential for future research in analyzing learners' emotions and enhancing learning approaches within mulsemmedia learning environments. Some of the key areas for exploration are listed below:

- **Emotion Recognition and Mulsemmedia:** Emotion recognition combined with mulsemmedia learning holds the potential to create highly personalized and immersive educational experiences. By understanding a learner's emotional state, a mulsemmedia system can dynamically adjust the sensory stimuli to optimize learning outcomes. For example, if a student is detected to be frustrated, the system might reduce the complexity of a task or provide soothing sensory inputs to alleviate stress.
- **Creating New FER Databases:** Developing various new facial expression recognition databases that specifically include facial expressions related to learning activities can enhance the accuracy and applicability of FER systems in educational settings.
- **Adaptive Learning Systems:** Generative AI (GEN AI) models can help in creating adaptive learning systems that adjust content and delivery based on the learner's emotional state, thus improving engagement and effectiveness. These systems can personalize learning paths and provide targeted feedback to meet individual learner needs.
- **Composite Facial Expression Research:** Composite facial expression recognition has significant investigational promise, but there are only a few studies on it, and their

results are not outstanding. Further research in this area is needed to improve the identification and interpretation of composite facial expressions, which can provide deeper insights into learners' emotional states.

- **Immersive Learning Content:** Introducing mulsemmedia content grounded in immersive learning approaches, such as mixed reality and the metaverse, has the potential to revolutionize education by offering highly engaging and realistic learning experiences. These technologies can create virtual environments that simulate real-world scenarios, making the learning process more interactive and effective.