SPECIMEN FORMAT FOR THESES OF MONTH

Faculty	**:**	School of Physical Science and Computational Sciences
Department	**:**	Computer Science
Branch/ Area:	**:**	Deep Learning
Sub Subject Heading:	**:**	Deep Learning
Candidate's Name	**:**	**Mohana. M**
Candidate's Address with email	**:**	577, 7th ward, Kandhan Nagar, Udaiyampalayam, Modachur, Gobi – 638476 mohanamurukan@gmail.com Ph: 9677574864
Title of the thesis	**:**	Deep Learning-Based Facial Expression Recognition for Analysing Learner Engagement in Mulsemedia Enhanced Teaching
(i) In Roman Script (ii) In roman Script		
Nomenclature of Degree:	**:**	**PhD in Computer Science**
Month & Year of Enrolment:	**:**	**January – 2021**
Month & Year of Registration:	**:**	**January -2021**
Month & Year of Submission:	**:**	**November – 2024**
Month & Year of Award	**:**	**April – 2025**
Name of Supervisor	**:**	**Dr. P. Subashini**
Designation of Supervisor	**:**	**Professor**
Centre/department/school in Which research was conducted	**:**	Centre for Machine Learning and Intelligence, Department of Computer Science

University's Name & Address	**:**	Avinashilingam Institute For Home Science and Higher Education for Women Coimbatore – 641 043, Tamilnadu, India.

Abstract within 300 words:

In the current digital era, technology-enhanced learning is evolving rapidly, setting new trends in educational environments and enabling students to learn more efficiently than ever before. Conventional learning course content typically involves only two sensory modalities—audio and video—which limits its ability to engage learners fully. In contrast, immersive learning course content and environments incorporate multiple senses, allowing learners to interact with multimedia content in ways that go beyond sight and sound. This approach, known as mulsemedia (multiple sensorial media), posits that engaging sensory channels—such as audio, visual, haptic, olfactory, temperature, gustatory, and even airflow— can significantly reinforce the learning process. Furthermore, measuring learner engagement is essential to ensuring that learners remain actively involved in learning. Various detection methods can assess engagement levels; in this study, we focus on analyzing engagement through facial expressions, particularly in a mulsemedia-synchronized learning environment.

Modern Facial Expression Recognition (FER) systems have achieved significant results through deep learning techniques. However, existing FER systems face two primary challenges: overfitting due to limited training datasets, and additional complications unrelated to expressions, such as occlusion, pose variations, and illumination changes. To improve the performance of FER in analyzing learners' engagement within a mulsemedia-based learning environment and to address some of these challenges, we propose three key aspects.

In our first study, face detection is a crucial step for identifying and cropping faces to train FER models. We observed that the conventional Viola-Jones face detection algorithm often produced false positives, particularly in complex images containing multiple faces or cluttered backgrounds. To address this issue, we enhanced the Viola-Jones algorithm by integrating particle swarm optimization to improve prediction accuracy in challenging images. The integration optimizes threshold selection and refines feature selection, enabling AdaBoost within the Viola-Jones framework to focus on the most relevant features for constructing a robust classifier. This enhancement significantly reduces false positives by fine-tuning feature selection and cascade thresholds, thereby improving prediction accuracy in complex environments.

In our second study, we observed that existing supervised FER approaches are inadequate for analyzing spatiotemporal features in real-time environments involving dynamic facial movements. To overcome this limitation, we introduced a fusion of convolutional neural networks and Bidirectional Long Short-Term Memory (Bi-LSTM) networks to recognize emotions from facial expressions and capture relationships between sequences of expressions. Our approach employs a VGG-19 architecture with optimized hyperparameters and TimeDistributed layers to independently extract spatial features

from each frame within a sequence. These spatial features are subsequently fed into a Bi-LSTM, which captures temporal relationships across frames in both forward and backward directions. This fusion enhances the model's ability to recognize emotions from expression sequences. The proposed method achieves significant accuracy in FER analysis, with results compared against baseline techniques.

In our third study, we introduced a Deep Semi-Supervised Convolutional Sparse Autoencoder to address the limitations of supervised FER approaches, particularly their reliance on extensive datasets and the challenges posed by imbalanced facial expression distributions, which can adversely affect model performance. This approach consists of two main stages. In the first stage, a deep convolutional sparse autoencoder is trained on unlabeled facial expression samples. Sparsity is introduced in the convolutional block through penalty terms, encouraging the model to focus on extracting the most relevant features for latent space representation. In the second stage, the trained encoder's feature map is connected to a fully connected layer with a softmax activation function for fine-tuning, forming a semisupervised learning framework. This approach enhances FER accuracy in real-time environments. Furthermore, these two approaches were conducted using the Extended Cohn-Kanade+, Japanese Female Facial Expression, and an In-house dataset. Model performances are evaluated using metrics including accuracy, precision, recall, F1-score, the confusion matrix, and the receiver operating characteristic curve.

Finally, all proposed methods were integrated to effectively analyze learner engagement levels in mulsemedia-synchronized learning environments. To achieve this, a mulsemedia-synchronized web portal was developed, incorporating olfactory, vibration, and airflow effects. The FER system mapped eight facial expressions to three engagement levels—highly engaged, engaged, and disengaged—based on the system's predicted probability scores and predefined threshold values. The final results demonstrate that mulsemedia-based learning significantly improved learning outcomes and memory retention compared to conventional methods.

i) Major objectives:

Primary Objective:

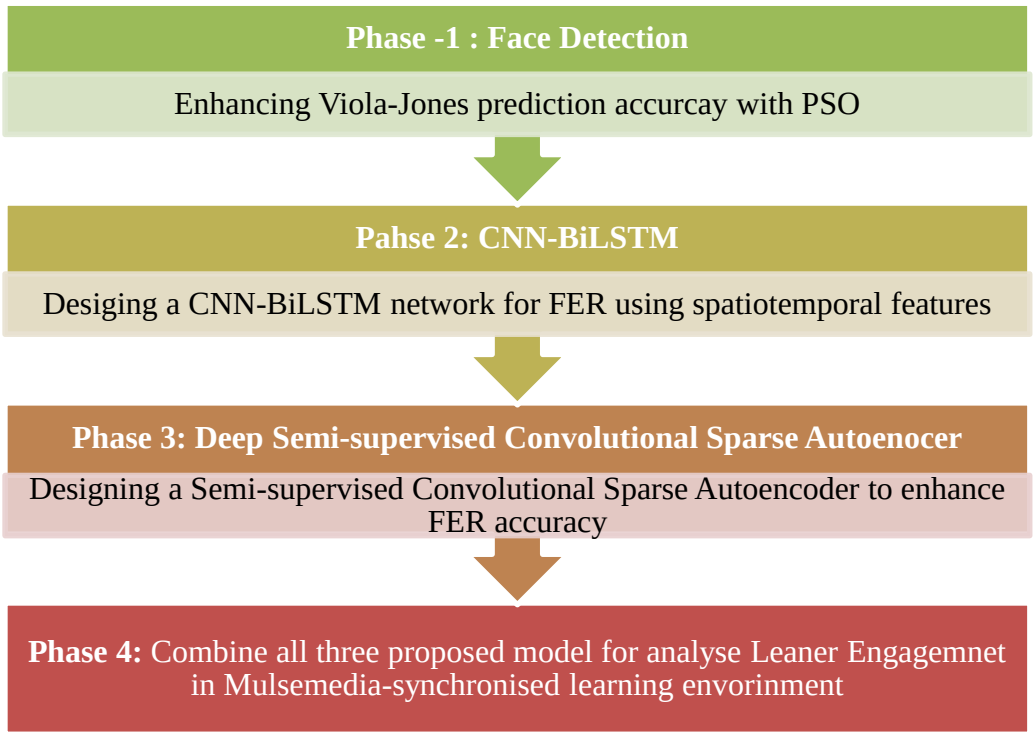
- To develop an effective FER model for predicting learners' engagement through universal expressions using a deep learning approach in a mulsemedia-enhanced learning environment.

Secondary objectives:

- To experiment with various deep learning approaches (3D-CNN, LSTM, and various types of Autoencoder) for building effective way of FER systems in Mulsemedia learning environments using a universal facial expression dataset.

- To investigate the effectiveness of learners' satisfaction with mulsemediasynchronised content, fostering improved engagement and knowledge retention among learners.
- Maximize classification accuracy and detection rate in FER.
- Minimize prediction time for FER.
- Maximize precision and recall scores in emotion classification.

ii) Methodology:



iii) Findings:

The thesis makes five contributions by exploring various aspects of FER, categorized as follows:

1. **Enhancing Viola-Jones Face Detection Algorithm Prediction Accuracy**: The conventional Viola-Jones algorithm employs AdaBoost for classifying faces in images and videos. The challenge lies in working with cluttered real-time facial images. AdaBoost needs to search through all possible thresholds for all samples to find the minimum training error when receiving features from Haar-like detectors. This exhaustive search consumes significant time to discover the best threshold values and optimize feature selection to build an efficient classifier for face detection. To address this, we proposed enhancing the conventional Viola-Jones algorithm by incorporating Particle Swarm Optimization (PSO) to improve its predictive accuracy, particularly in complex face images. We leverage PSO in two key areas within the Viola-Jones framework. Firstly, PSO is employed to dynamically select optimal threshold

values for feature selection, thereby improving computational efficiency. Secondly, we adapt the feature selection process using AdaBoost within the Viola-Jones algorithm, integrating PSO to identify the most discriminative features for constructing a robust classifier. Our approach significantly reduces the feature selection process time and search complexity compared to the traditional algorithm, particularly in challenging environments. We evaluated our proposed method on a comprehensive face detection benchmark dataset, achieving impressive results, including an average true positive rate of 98.73% and a 2.1% higher average prediction accuracy compared to both the conventional Viola-Jones approach and contemporary state-of-the-art methods.

2. **FER using CNN-BiLSTM architecture**: Existing FER systems primarily focus on spatial features for emotion recognition, but they struggle to accurately identify emotions from dynamic sequences of facial expressions in real-time. We propose deep learning techniques that fuse CNN and Bidirectional LSTM (BiLSTM) to recognize emotions by leveraging spatiotemporal features, enabling the identification of relationships between sequences of facial expressions. This approach employs a hyperparameter-tuned VGG-19 model with time-distributed layers to automatically extract spatial features from a sequence of images, addressing the limitations of conventional feature extraction methods. Next, these feature sequences are fed into a BiLSTM network to analyze temporal features in both directions, enabling the recognition of emotions from a sequence of expressions. Experimental results demonstrate that the proposed method outperforms both baseline methods and state of-the-art approaches.

3. **FER using Semi-supervised Convolutional Sparse Autoencoder:** Most deep learning approaches in supervised FER systems heavily rely on large, labelled datasets. Implementing FER in CNNs often requires many layers, leading to extended training times and difficulties in finding optimal parameters. This can result in challenges in creating distinct facial expression patterns for classification, leading to poor real-time emotion classification. Therefore, we introduce a new approach called the Deep Semi-supervised Convolutional Sparse Autoencoder (DSCSA) to address these issues and enhance FER performance and prediction accuracy. This approach comprises two parts: Initially, a deep convolutional sparse autoencoder is trained with unlabeled facial expression samples. Here, sparsity is introduced in the convolutional block to enforce penalties, focusing on more relevant features for feature representation in the latent space. A trained encoder with a feature map is connected to a fully connected layer with softmax for final fine-tuning with learned weights and labeled facial expression samples in a semi-supervised approach for emotion classification. The results were analyzed using established state-of-the-art techniques.

4. **Designing a Mulsemedia-enabled Web portal and analysis of learner engagement:** We designed an IoT-based mulsemedia-synchronized learning platform that uses affordable components, such as cooling fans, humidifiers, and haptics, to create a multisensory learning environment. This system provides learners with an immersive experience by incorporating sensory effects, such as the aroma of rosemary, along with vibrotactile and airflow effects associated with thunder and lightning. These effects are synchronized with traditional audiovisual content. The study involved 70 participants pursuing science degrees, divided into two equal-sized groups: an experimental group and a control group, to analyze the impact of

mulsemedia on their learning experience. The results showed that mulsemedia-based learning significantly improved learning outcomes and increased enjoyment levels. Additionally, it enhanced the sense of reality in the conventional learning environment. Finally, the proposed FER model was integrated to analyze learner engagement levels in the mulsemedia-enhanced learning environment.

Examiners

Internal Examiner: Dr.(Mrs.) R. Srivaramangai

Head, Department of Information Technology,

University of Mumbai

Mumbai-400098

External Examiner: Dr.Manoranjan Paul

Professor

School of Computing, Mathematics and Engineering

Charles Sturt University

Australia