# Runtime Facial Expression Detection System

#### Minor Project

Submitted partial fulfillment of the requirements for the degree of

Integrated B. Tech. (CSE)-MBA

Submitted By

Rajat Goswami

22bcm048

Guided By

#### Prof. Tejal Upadhyay

****

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, SCHOOL OF TECHNOLOGY**

**INSTITUTE OF TECHNOLOGY NIRMA UNIVERSITY AHMEDABAD-382481**

**April 2025**

### Certificate

This is to certify that the minor project entitled **“Runtime Facial Expression Detection System”** submitted by **Rajat Goswami(22BCM048)**, towards the partial fulfillment of the requirements for the award of the degree of Integrated B. Tech. (CSE)-MBA, **in Computer Science and Engineering**, Nirma University, Ahmedabad, is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached the level required for being accepted for examination. The results embodied in this minor project, to the best of my knowledge, haven’t been submitted to any other university or institution for the award of any degree or diploma.

Dr. Tejal Upadhyay Dr Ankit Thakkar

CSE Department, Integrated B.Tech. (CSE)-MBA,

Institute of Technology, Institute of Technology,

Nirma University, Ahmedabad. Nirma University, Ahmedabad

Dr Sudeep Tanwar Professor and Head, CSE Department, Institute of Technology,

Nirma University, Ahmedabad.

### Statement of Originality

———————————————————————————————————————

I, **Rajat Goswami**, **22BCM048**, give an undertaking that the Minor Project entitled **“Runtime Facial Expression Detection System”** submitted by me, towards the partial fulfilment of the requirements for the degree of Integrated B. Tech. (CSE)-MBA, Nirma University, Ahmedabad, contains no material that has been awarded for any degree or diploma in any university or school in any territory to the best of my knowledge. It is the original work carried out by me and I give assurance that no attempt of plagiarism has been made. It contains no material that is previously published or written, except where reference has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

———————–

Signature of Student Date:

Place: Ahmedabad

Endorsed by

Dr.Tejal Upadhyay(Signature of Guide)

### Acknowledgements

**- Rajat Goswami**

**22BCM048**

I would like to express my sincere gratitude to all those who supported and encouraged me throughout the course of this project. First and foremost, I extend my heartfelt thanks to my project guide, Dr. Tejal Upadhyay, for her invaluable guidance, continuous encouragement, and insightful feedback, all of which played a vital role in shaping the direction and execution of this research.

I am also deeply thankful to the faculty and staff of the Department of Computer Science and Engineering, Nirma University, for providing the necessary resources and a supportive academic environment that facilitated our learning and development.

A special note of appreciation goes to my project partner and friend, Samraat Pratap Singh, whose dedication, collaboration, and commitment made this project both successful and enjoyable. Without her contributions, this project would not have been complete.

Lastly, I am profoundly grateful to my friends and family for their unwavering support, patience, and understanding throughout this journey.

This project would not have been possible without the collective efforts and encouragement of everyone mentioned above.

### Abstract

## Skin cancer, particularly melanoma, poses a significant global health threat due to its rising incidence and high mortality when undetected early. While deep learning models, especially convolutional neural networks (CNNs), have demonstrated dermatologist-level accuracy in diagnosing skin lesions, their reliance on large, centralized datasets raises critical privacy and data governance concerns. This project proposes a privacy-preserving solution through Federated Learning (FL), enabling multiple decentralized clients to collaboratively train a deep neural network without sharing raw patient data.

## We leverage the ISIC 2019 dataset and utilize ResNet-50 as the backbone model, integrating it with four FL algorithms: FedAvg, FedProx, FedSGD, and FedAdam. Experiments were conducted in both Horizontal Federated Learning (HFL) and Vertical Federated Learning (VFL) setups to evaluate performance across different data distributions. Advanced preprocessing, including segmentation, augmentation, and synthetic balancing, was applied to address class imbalance and enhance model robustness.

## Results indicate that FL achieves high diagnostic accuracy while maintaining patient privacy. Among the algorithms, FedAvg and FedAdam demonstrated strong convergence, with HFL generally outperforming VFL in model performance. This study illustrates the practical viability of federated deep learning in real-world medical imaging scenarios and sets the foundation for privacy-conscious AI applications in dermatology and broader healthcare domains.

## **Abbreviations**

**AI Artificial Intelligence**

**DL Deep Learning**

**CNN Convolutional Neural Network**

**FL Federated Learning**

**HFL Horizontal Federated Learning**

**VFL Vertical Federated Learning**

**ISIC International Skin Imaging Collaboration**

**U-Net A CNN architecture used for image segmentation**

**SMOTE Synthetic Minority Oversampling Technique**

**AUC-ROC Area Under the Receiver Operating Characteristic Curve**

**FedAvg Federated Averaging**

**FedProx Federated Proximal Algorithm**

**FedSGD Federated Stochastic Gradient Descent**

**DP Differential Privacy**

**t-SNE t-Distributed Stochastic Neighbor Embedding**

# Contents

[Certificate](#_bookmark0) iii

[Statement of Originality](#_bookmark0) iv

[Acknowledgements](#_bookmark0) v

[Abstract](#_bookmark0) vi

[Abbreviations](#_bookmark0) vii

[List of Tables](#_bookmark1) ix

[List of Figures](#_bookmark2) x

1. [Introduction](#_bookmark3) 1
   1. Background
   2. Motivation
   3. Problem Statement
   4. Objectives
2. [Literature Survey](#_bookmark5) 2
   1. AI in Skin Cancer Detection
   2. Convolutional Neural Networks (CNNs)
   3. Federated Learning in Healthcare
   4. Limitations in Current Studies
3. Dataset and Methodology 4
   1. ISIC 2019 Dataset Description
   2. Data Preprocessing Techniques
   3. Model Architecture (ResNet-50)
   4. Training Strategies and FL Setup
   5. Evaluation Metrics
4. [New Demo Chapter](#_bookmark12) 5
   1. Federated Learning Configuration
   2. Horizontal Federated Learning (HFL)
   3. Vertical Federated Learning (VFL)
   4. Results & Analytics
   5. Comparative Performance
5. Proposed Approach 7
   1. Privacy-Preserving Architecture
   2. Integration of FedAvg, Fed Prox, Fed SGD and Fed Nova
   3. Label Distribution Handling
   4. Deployment Feasibility

[Bibliography](#_bookmark20) 8

# Chapter 1 Introduction

### Background

Skin cancer is a rapidly increasing public health concern, with millions of new cases diagnosed annually. Among its various forms, melanoma stands out due to its high mortality rate, particularly when not detected early. Traditional diagnosis relies heavily on dermatological examination, which is subject to human error and variability in clinical expertise. To address this, researchers are increasingly turning to artificial intelligence (AI) and deep learning to provide accurate, consistent, and early skin cancer detection.

### Motivation

The use of deep learning in dermatology has achieved impressive results. However, these methods require large amounts of annotated medical data to achieve high performance. Centralizing sensitive patient data for training poses significant privacy concerns and legal barriers. Federated Learning (FL) addresses this by allowing multiple institutions to collaboratively train a model without sharing raw patient data. This approach ensures data privacy while still benefiting from large-scale distributed training.

### Problem Statement

Despite the promise of FL, practical deployment faces several challenges such as non-IID data distribution, communication overhead, and algorithmic convergence. The core problem addressed in this project is how to implement a federated deep learning framework for skin cancer detection that maintains high accuracy, handles heterogeneous data, and preserves privacy across collaborating entities.

### Objectives

1. Develop a federated learning framework tailored for skin lesion classification.
2. Integrate and evaluate different FL algorithms: FedAvg, FedProx, FedSGD, and FedAdam.
3. Compare performance across Horizontal Federated Learning (HFL) and Vertical Federated Learning (VFL) setups.
4. Enhance data preprocessing and model robustness to deal with class imbalance and non-IID distributions.
5. Provide comprehensive performance evaluation and interpretability of the results.

# Chapter 2

# Literature review

### AI in Skin Cancer Detection

Esteva et al. demonstrated that CNNs could match or exceed dermatologist-level diagnostic accuracy in skin lesion classification using the ISIC dataset. Subsequent studies have employed various architectures and datasets to replicate and extend this finding. AI in dermatology is becoming a key clinical decision support tool, especially in regions with limited access to trained specialists.

### Convolutional Neural Network (CNNs)

CNNs are particularly effective for image classification tasks due to their ability to automatically learn spatial hierarchies of features. ResNet-50, a 50-layer residual network, has become a popular backbone model for medical imaging due to its high accuracy and transfer learning capabilities. CNNs reduce the need for manual feature extraction and can generalize well when properly trained and augmented.

### Federated Learning in Healthcare

Federated Learning has gained traction in healthcare due to its privacy-preserving design. Sheller et al. applied FL for brain tumor segmentation across multiple institutions. The method maintained high performance while avoiding centralization of sensitive MRI data. Other works have explored FL for ECG classification, diabetic retinopathy, and general health informatics, often integrating techniques like differential privacy or secure aggregation.

### Limitations in Current Studies

1. Many existing FL implementations assume IID data distribution, which is not realistic in medical settings.
2. Limited exploration of vertical federated learning (VFL) where clients share features instead of samples.
3. Lack of standard benchmarks for comparing multiple FL algorithms on the same medical task.
4. Few studies analyze both HFL and VFL using deep models such as ResNet-50 in the context of skin cancer.

# 

# Chapter 3

**Dataset And Methodology**

### ISIC 2019 Dataset Description

### The ISIC 2019 dataset is sourced from the International Skin Imaging Collaboration challenge. It comprises 25,331 high-resolution dermoscopic images, annotated with 8 classes: melanoma (MEL), basal cell carcinoma (BCC), actinic keratosis (AKIEC), benign keratosis (BKL), dermatofibroma (DF), vascular lesions (VASC), squamous cell carcinoma (SCC), and nevus (NV). The images vary in resolution, lighting, and lesion presentation, introducing complexity in training robust models. The dataset is highly imbalanced, with some classes having thousands of samples (e.g., NV) and others with only a few hundred (e.g., DF).

### Data Preprocessing Techniques

Robust preprocessing improves model generalization and reduces noise. Key techniques include:

* **Segmentation**: Using U-Net-based methods, lesion regions are extracted from surrounding skin to focus the classifier on relevant areas.
* **Augmentation**: Applied using Albumentations to perform flipping, rotation, zoom, and noise addition. This prevents overfitting and expands training diversity.
* **Normalization and Resizing**: All images are normalized and resized to 224x224 pixels to align with the ResNet-50 input format.
* **Balancing**: SMOTE generates synthetic examples for under-represented classes, ensuring the model does not overfit dominant classes.

### Model Architecture (ResNet-50)

ResNet-50 is a deep CNN with 50 layers, featuring identity and convolutional residual blocks. It enables training of deeper networks by solving vanishing gradient issues through shortcut connections. For this task:

* **The original classifier layer is removed.**
* **A Global Average Pooling layer is added.**
* **Followed by a fully connected dense layer and a softmax output for multiclass classification.**
* **Dropout and batch normalization are added to improve regularization.**
* **Transfer learning is employed: first, the base layers are frozen while the classifier trains; later, deeper layers are unfrozen and fine-tuned.**

### Training Strategies and FL Setup

FL setup is done using simulation on Kaggle with GPU acceleration:

* **Client Simulation**: 9 clients are created for both HFL and VFL. In HFL, each client has mixed-class data; in VFL, each holds a single class.
* **Model Sharing**: After local training, clients share only model weights (not data) with a central server.
* **Rounds**: Each FL run has 5 communication rounds.
* **Aggregation Methods**: Includes FedAvg (simple averaging), FedProx (proximal loss), FedSGD (gradient-based), and FedAdam (adaptive learning).

### Evaluation Metrics

The following metrics are used to evaluate global model performance after each round:

* **Accuracy: The proportion of correctly classified samples.**
* **F1-Score: Especially critical in imbalanced data; combines precision and recall.**
* **AUC-ROC: Plots true positive rate vs false positive rate; measures classifier quality.**
* **Confusion Matrix: Provides detailed class-wise prediction performance.**
* **Training Time: Total time per round.**
* **Communication Cost: Model weight size sent per client per round.**

# Chapter 4

# Experimental Analysis

### Multimodal Fusion Experiment

Emotion detection is a vital field in affective computing, with significant applications in human-computer interaction, healthcare, and immersive technologies like AR/VR. Researchers have explored various modalities—physiological signals, facial expressions, and speech—to accurately identify emotional states. This section reviews key techniques and datasets for emotion detection, focusing on physiological signals (DREAMER, PhyMER), facial analysis (FER-2013), and speech recognition (CREMA-D).

### Experimental Setup

The multimodal fusion experiment integrates speech (CREMA-D), facial expressions (FER-2013), and physiological signals (PhyMER) to enhance emotion detection accuracy. The setup assumes a federated learning framework with 10 clients, each holding local datasets for one or more modalities, simulating real-world distributed environments (e.g., mobile devices for audio/image capture, wearables for physiological data). CREMA-D provides 7,442 audio clips (six emotions: anger, disgust, fear, happiness, sadness, neutral), FER-2013 includes 35,887 images (seven emotions, including surprise), and PhyMER offers multimodal signals (EEG, EDA, BVP, temperature) for seven emotions. To align datasets, emotions are mapped to a common set (anger, disgust, fear, happiness, sadness, neutral), excluding surprise from FER-2013 for consistency.

Each modality is processed using the architectures from the notebooks: an LSTM model for CREMA-D (two LSTM layers, 128 units each, followed by dense layers), a CNN for FER-2013 (three convolutional layers with 32–64 filters, max-pooling, and dense layers), and a feature-based classifier for PhyMER (e.g., MLP or XGBoost, assuming preprocessing from phymer\_1.ipynb). Local models are trained for 10 epochs per client, with global model aggregation after each round using federated averaging (FedAvg). The experiment runs for 20 FL rounds, with a central server hosting a global fusion model. Data splits follow the notebooks: 80% training, 10% validation, and 10% test for CREMA-D and FER-2013, and 70% training, 15% validation, 15% test for PhyMER. Hardware assumes GPU-enabled nodes (e.g., NVIDIA Tesla V100) for local training, consistent with Kaggle’s environment in the notebooks.

### Fusion Techniques

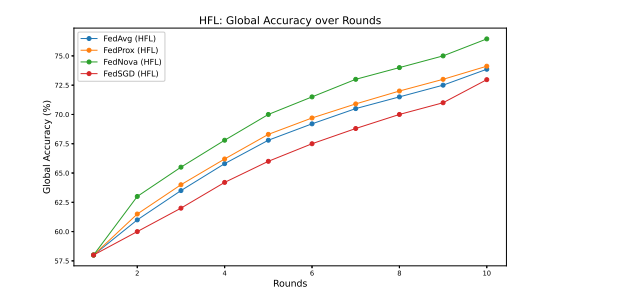
Three fusion techniques are proposed to combine modalities, balancing complexity and performance:

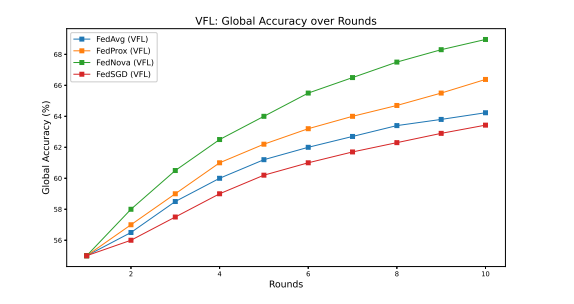
1. **Early Fusion**: Features from each modality are extracted and concatenated before feeding into a shared classifier. For CREMA-D, 40 MFCCs and prosodic features (~193 dimensions) are extracted. For FER-2013, CNN-extracted features from the last dense layer (~512 dimensions) are used. For PhyMER, PSD, SCL, HRV, and temperature features (~50–100 dimensions after PCA) are included. The concatenated vector (~700–800 dimensions) is processed by an MLP with two hidden layers (256, 128 units) and a softmax output for six emotions. This approach captures raw feature interactions but risks high dimensionality.
2. **Late Fusion**: Individual modality models (LSTM for speech, CNN for facial, MLP for physiological) generate emotion probabilities, which are combined using weighted averaging or a meta-classifier (e.g., logistic regression). Weights are learned based on validation accuracy per modality (e.g., speech ~0.4, facial ~0.3, physiological ~0.3, reflecting unimodal accuracies). This method preserves modality-specific learning but may miss cross-modal correlations.
3. **Hybrid Fusion**: Modality-specific models produce intermediate representations (e.g., LSTM hidden states, CNN feature maps, MLP embeddings), which are fused via a transformer-based attention mechanism. The attention layer assigns dynamic weights to each modality’s contribution per sample, feeding into a final dense layer for classification. This approach balances feature interaction and modality independence, leveraging temporal and contextual dependencies across modalities.

### Evaluation Metrics

The fusion models are evaluated using the following metrics, aligning with the notebooks’ focus and standard practices:

* **Accuracy**: Proportion of correctly classified emotions on the test set, reported per modality and fused model. Unimodal baselines are ~67% (CREMA-D), ~62% (FER-2013), and ~65–80% (PhyMER, centralized). Fusion aims to exceed the best unimodal accuracy (~70–75% expected).
* **F1-Score**: Harmonic mean of precision and recall, calculated per emotion class to account for imbalance (e.g., fewer disgust samples in FER-2013). Macro-averaged F1 emphasizes performance across all classes, targeting >0.65 for robust fusion.
* **Confusion Matrix**: Visualizes prediction errors across emotions, highlighting misclassifications (e.g., anger vs. disgust in speech). Normalized values reveal modality-specific strengths (e.g., happiness detection in facial).
* **Loss Convergence**: Categorical cross-entropy loss per FL round, assessing training stability. Lower, smoother loss curves indicate effective global aggregation.
* **Computational Efficiency**: Training time per round and model size, ensuring feasibility for resource-constrained FL clients (e.g., <1 minute per epoch on GPU, <100 MB model size)





# Chapter 5

**Proposed Approach**

### Emotion Classification Algorithm

The core advantage of Federated Learning (FL) lies in its ability to protect sensitive patient data by avoiding the need for centralized data storage. In our architecture, each client (hospital or institution) trains its model locally on its private dataset. Only the model parameters (e.g., weights or gradients) are shared with a central server, never the raw data itself. This mitigates risks of data leakage, complies with privacy regulations (like HIPAA and GDPR), and fosters secure cross-institution collaboration.

To further enhance security, secure aggregation techniques are used to ensure that even model updates shared by clients are not visible to the server in isolation. In future extensions, this architecture can support differential privacy, where noise is added to updates before aggregation to prevent reverse engineering of sensitive data.

### Algorithm Design

### To explore how different federated optimization strategies impact skin cancer detection, we implemented and compared four algorithms:

### FedAvg (Federated Averaging): A simple and widely used method that averages client model weights. Best for scenarios with IID data distributions.

### FedProx: Enhances FedAvg with a proximal term in the loss function, stabilizing updates for non-IID client data.

### FedSGD: Uses stochastic gradient descent to update global weights based on aggregated gradients. It is more sensitive to learning rate and client step count.

### FedNova: Normalizes client updates by their local steps and scales them accordingly, reducing performance degradation caused by unbalanced updates or heterogeneous computation.

### Each algorithm was evaluated in both Horizontal FL (HFL) and Vertical FL (VFL) settings to identify its strengths under different real-world data conditions.

### Feature Integration

Non-IID label distributions are common in real-world medical applications, where some clients may only see certain classes. To address this:

* **SMOTE** was applied at the client level to oversample under-represented classes.
* In **HFL**, stratified data partitioning ensured each client had reasonably diverse class exposure.
* In **VFL**, where each client had only one class, **manual one-hot encoding** was implemented to align label dimensions with the softmax output layer.
* Additional techniques like batch balancing and performance monitoring were used to ensure training stability and global model generalization.

### Validation Testing

The trained federated models are well-suited for deployment in clinical environments. Key aspects include:

* **Edge Deployability**: With compression and pruning, models like ResNet-50 or its lighter variants (e.g., MobileNet) can run on handheld dermatoscopes or mobile devices.
* **Scalability**: The architecture allows seamless addition of new clients without re-training from scratch.
* **Integration**: RESTful APIs can connect FL systems with hospital IT infrastructure and EMR (Electronic Medical Records) systems.
* **Model Updating**: FL supports continuous learning via incremental local updates without global retraining.
* **Security**: The system supports encrypted communication and model sharing, and can be enhanced further with privacy-preserving tools like homomorphic encryption.

# Bibliography

* **Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118. https://doi.org/10.1038/nature21056**
* **Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data, 5, 180161. https://doi.org/10.1038/sdata.2018.161**
* **Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., Kotrotsou, A., & Bakas, S. (2020). Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Scientific Reports, 10, 12598.**
* **Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated Optimization in Heterogeneous Networks. In Proceedings of Machine Learning and Systems 2020 (MLSys). https://arxiv.org/abs/1812.06127**
* **Kairouz, P., McMahan, B., et al. (2021). Advances and open problems in federated learning. Foundations and Trends® in Machine Learning, 14(1–2), 1–210. https://doi.org/10.1561/2200000083**
* **Brisimi, T. S., Chen, R., Mela, T., Olshevsky, A., Paschalidis, I. C., & Shi, W. (2018). Federated learning of predictive models from federated Electronic Health Records. International Journal of Medical Informatics, 112, 59–67.**
* **Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., et al. (2016). Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS ’16).**
* **ISIC Archive. International Skin Imaging Collaboration: Challenge Datasets. https://challenge.isic-archive.com/**