| **Project Title** | **Tuberculosis Detection Using Deep Learning** |
| --- | --- |
| **Skills take away From This Project** | Python scripting, Deep Learning, Computer Vision, Transfer Learning, Model Evaluation, Streamlit, AWS Deployment |
| **Domain** | Healthcare, Medical Imaging, Deep Learning |

**Problem Statement**

Develop a deep learning-based model to classify chest X-ray images as either normal or showing signs of tuberculosis (TB). The system will preprocess and augment image data, train multiple deep learning models, and evaluate their performance. The final application will provide an interface for uploading X-ray images and receiving predictions, deployed via Streamlit on AWS.

**Business Use Cases**

1. **Early Detection of Tuberculosis:**
   * Aid radiologists and healthcare professionals in diagnosing TB from X-ray images.
2. **Automated Screening in Remote Areas:**
   * Provide medical imaging analysis where radiologists are not readily available.
3. **Reducing Diagnostic Errors:**
   * Support decision-making by providing AI-powered second opinions.
4. **Research and Analysis:**
   * Help researchers analyze TB trends and model effectiveness in detecting abnormalities.

**Approach**

1. **Data Preparation:**
   * Use the provided dataset containing normal and TB X-ray images.
   * Split the dataset into training, validation, and test sets.
2. **Data Cleaning and Preprocessing:**
   * Resize images, normalize pixel values, and apply augmentations.
   * Ensure class balance and handle missing/corrupt images.
3. **Exploratory Data Analysis (EDA):**
   * Visualize image distribution, pixel intensity statistics, and class balance.
4. **Model Development:**
   * Train multiple deep learning models using Transfer Learning (ResNet50, VGG16, EfficientNetB0).
   * Experiment with hyperparameter tuning, dropout, and data augmentation.
5. **Model Evaluation and Comparison:**
   * Use metrics like accuracy, precision, recall, F1-score, and ROC-AUC.
   * Compare model performance and select the best-performing model.
6. **Application Development:**
   * Build a Streamlit-based interface for users to upload X-ray images and receive predictions.
7. **Deployment:**
   * Deploy the trained model on AWS for scalability and accessibility.

**Data Flow and Architecture**

1. **Data Collection & Preprocessing:**
   * Load X-ray images and preprocess them using OpenCV and TensorFlow.
2. **Processing Pipeline:**
   * Implement data augmentation, feature extraction, and normalization.
3. **Model Training:**
   * Train CNN models using TensorFlow/Keras.
   * Save trained models for deployment.
4. **Deployment:**
   * Create a Streamlit-based front-end for users to interact with the model.
   * Host the application on AWS EC2 or Elastic Beanstalk.

**Dataset:**

[**tuberculosis-chest-x-rays-images**](https://www.kaggle.com/datasets/yasserhessein/tuberculosis-chest-x-rays-images)

**Dataset Explanation:**

The dataset contains a total of 3008 chest X-ray images, categorized into two groups:

1- Tuberculosis (TB) Patients:

* Number of Images: 2494
* Description: These images represent chest X-rays from patients diagnosed with tuberculosis (TB).

2- Normal Patients:

* Number of Images: 514
* Description: These images represent chest X-rays from individuals with no diagnosed abnormalities.

**Results**

By the end of this project, learners will achieve:

* A preprocessed and augmented dataset ready for deep learning.
* Multiple CNN-based models trained and evaluated.
* A deployed TB detection system accessible via a Streamlit interface.
* A fully functional application hosted on AWS.

**Project Evaluation Metrics:**

1. **Data Preprocessing Quality:**
   * How well the data is cleaned, augmented, and structured.
2. **Model Performance:**
   * Accuracy, precision, recall, F1-score, and ROC-AUC.
3. **Application Functionality:**
   * User-friendly Streamlit interface for predictions.
4. **Deployment Quality:**
   * Accessibility and scalability of the AWS-hosted application.

**Technical Tags:**  
 Python, Deep Learning, Computer Vision, CNN, Transfer Learning, Streamlit, AWS Deployment, TensorFlow, Image Processing

**Deliverables:**

1. **Data Preparation:**
   * Evaluate how well the dataset is cleaned and preprocessed.
2. **Model Training:**
   * Assess performance across different architectures.
3. **Evaluation Metrics:**
   * Compare models using classification metrics.
4. **Application Deployment:**
   * Ensure the AWS-hosted interface is accessible and responsive.

**Timeline:**

* 2 Weeks

**References:**

| Project Live Evaluation Metrics | [Project Live Evaluation](https://docs.google.com/document/u/0/d/1QisLD2kqDWFZJG2oDknKn2eMGi-Xq8oFPgA7UWSbcIQ/edit) |
| --- | --- |
| EDA Guide | [Exploratory Data Analysis (EDA) Guide](https://docs.google.com/document/d/1tHiTU1X9UwXSLySpJ-FVCohlf_8xpXwa75vlK9S6wl8/edit?usp=sharing) |
| Capstone Explanation Guideline | [Capstone Explanation Guideline](https://docs.google.com/document/d/1gbhLvJYY7J73lu1g9c6C9LRJvYemiDOdRDAEMe632w8/edit) |
| GitHub Reference | [How to Use GitHub.pptx](https://docs.google.com/presentation/d/1XHCbgUOqbcXNUyQ87vTlKdKRgAbBxtkA/edit?usp=sharing&ouid=109735616107417446342&rtpof=true&sd=true) |
| AWS recordings | [AWS](https://docs.google.com/document/d/1R9F44Pc9n5uLnGR9gtNQ65d16CelfkCbjfAbWaaVMog/edit?tab=t.0) |
| Streamlit recordings (English) | [Special session for STREAMLIT(11/08/2024)](https://docs.google.com/document/d/1aR3pUZFlCi8gicpF6aPHPESeFdOtGMlfob5PckresZk/edit?usp=sharing) |
| Streamlit recordings (Tamil) |  |
| Streamlit documentation | [Install Streamlit](https://docs.streamlit.io/get-started/installation) |
|  |  |
|  |  |

**PROJECT DOUBT CLARIFICATION SESSION ( PROJECT AND CLASS DOUBTS)**

**About Session:** The Project Doubt Clarification Session is a helpful resource for resolving questions and concerns about projects and class topics. It provides support in understanding project requirements, addressing code issues, and clarifying class concepts. The session aims to enhance comprehension and provide guidance to overcome challenges effectively.

**Note: Book the slot at least before 12:00 Pm on the same day**

**Timing: Monday-Saturday (4:00PM to 5:00PM)**

**Booking link :**[**https://forms.gle/XC553oSbMJ2Gcfug9**](https://forms.gle/XC553oSbMJ2Gcfug9)

**LIVE EVALUATION SESSION (CAPSTONE AND FINAL PROJECT)**

**About Session:** The Live Evaluation Session for Capstone and Final Projects allows participants to showcase their projects and receive real-time feedback for improvement. It assesses project quality and provides an opportunity for discussion and evaluation.

**Note: This form will Open only on Saturday (after 2 PM ) and Sunday on Every Week**

**Timing:**

**For DS and AIML**

**Monday-Saturday (05:30PM to 07:00PM)**

**Booking link :** [**https://forms.gle/1m2Gsro41fLtZurRA**](https://forms.gle/1m2Gsro41fLtZurRA)

**Evaluation Metrics** : [Project Live Evaluation](https://docs.google.com/document/d/1QisLD2kqDWFZJG2oDknKn2eMGi-Xq8oFPgA7UWSbcIQ/edit?usp=sharing)

| **Project Created By** | **Verified By** | **Approved By** |
| --- | --- | --- |
| [**Vinodhini**](mailto:vinodhini@guvi.in) | [**Nehlath Harmain**](mailto:nehlath@guvi.in) | [**Nehlath Harmain**](mailto:nehlath@guvi.in) |