CS400M MINOR PROJECT

A Project Report

Submitted in partial fulfilment of the requirements for the award of the Degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

BY

Sampath Kumar Midde, Chandani Bihani

(Btech/25139/22, Btech/25124/22)

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

BIRLA INSTITUTE OF TECHNOLOGY, MESRA

OFF-CAMPUS, JAIPUR

MO-2025

APPROVAL OF THE GUIDE

Recommended that the report entitled "**Deep Learning on Facial Depth Maps for Predicting Obstructive Sleep Apnea**" presented by Chandani Bihani (BTECH/25124/22) and Sampath Kumar Midde (BTECH/25139/22) under my supervision and guidance be accepted as fulfilling this part of the requirements for the award of the Degree of Bachelor of Technology (B.Tech.) in Computer Science and Engineering. To the best of my knowledge, the content of this report did not form a basis for the award of any previous degree to anyone else

Date:14/11/2025

Dr. Vishwambhar Pathak
Assistant professior
Dept. of CSE
BIT Mesra, Ranchi,
Off Campus Jaipur

DECLARATION CERTIFICATE

We certify that

a) The work contained in the report is original and has been done by ourselves under the general supervision of our supervisor.

b) The work has not been submitted to any other Institute for any other degree or diploma.

c) We have followed the guidelines provided by the Institute in writing the report.

d) We have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

e) Whenever we have used materials (data, theoretical analysis, and text) from other sources, we have given due credit to them by citing them in the text of the thesis and giving their details in the references.

f) Whenever we have quoted written materials from other sources, we have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

Sampath Kumar midde

Btech/25139/22

Chandani Bihani

Btech/25124/22

CERTIFICATE OF APPROVAL

This is to certify that the work embodied in this report entitled **“Deep Learning on Facial Depth Maps for Predicting Obstructive Sleep Apnea”**, is carried out by **Chandani Bihani (Roll No: BTECH/25124/22)** and **Sampath Kumar Midde (Roll No: BTECH/25139/22)** has been approved for the Degree of Bachelor of Technology (B.Tech.) in Computer Science and Engineering of Birla Institute of Technology, Mesra, Ranchi – Jaipur Off Campus

Date: 14/11/2025

Place: Jaipur campus

Internal Examiner **External Examiner**

(Chairman)

Incharge, Department of CSE

Abstract

Obstructive Sleep Apnea (OSA) is a common yet underdiagnosed condition that causes repeated airway blockages during sleep. Traditional diagnosis using Polysomnography (PSG) is accurate but costly and not easily accessible. This project explores the use of deep learning to predict OSA risk from lateral cephalometric X-ray images, offering a low-cost, non-invasive screening alternative.We implemented a convolutional neural network (CNN) based on ResNet-18, trained on the Aariz Cephalometric Dataset, which consists of grayscale craniofacial radiographs. Due to the lack of actual OSA diagnostic labels in the dataset, we used simulated binary labels to validate the pipeline structure. The images were resized to 224×224 pixels, normalized, and augmented for training.To make the model explainable, we integrated Grad-CAM**,** which highlighted facial regions that influenced predictions. The model consistently focused on clinically relevant zones like the jawline and upper airway. On the validation set of 147 images, the model achieved an accuracy of 48.98%**,** recall of 87.84%**,** and an F1-score of 63.41%**.**

Although trained with synthetic labels, the system demonstrates technical feasibility and lays the foundation for future integration with real OSA clinical data. This project provides a reproducible, scalable framework for AI-assisted OSA risk screening, with potential applications in mobile health and telemedicine.

ACKNOWLEDGEMENT

We would like to express our profound gratitude to our project guide **Dr. Vishwambhar Pathak** for his guidance and support during our thesis work. We benefited greatly by working under his guidance. It was his effort for which we are able to develop a detailed insight on this subject and special interest to study further. His encouragement motivation and support has been invaluable throughout our studies at BIT, Mesra, Jaipur Campus.

We convey our sincere gratitude to **Dr. Shripal Vijayvargiy**, Dept. of CSE, BIT, Mesra Jaipur Campus for providing us various facilities needed to complete our project work. We would also like to thank all the faculty members of CSE department who have directly or indirectly helped during the course of the study. We would also like to thank all the staff (technical and non-technical) and our friends at BIT, Mesra, Jaipur Campus who have helped us greatly during the course.

Finally, we must express our very profound gratitude to our parents for providing us with unfailing support and continuous encouragement throughout the years of our study. This accomplishment would not have been possible without them.

Our apologies and heartfelt gratitude to all who have assisted us yet have not been acknowledged by name.

Thank you.

DATE:

Sampath kumar midde
(Btech/25139/22)

Chandani Bihani

(Btech/25124/22)

CONTENTS

List of Figures & Tables

1. INTRODUCTION

1.1 Purpose

Obstructive Sleep Apnea (OSA) is a common yet frequently underdiagnosed condition characterized by repetitive airway obstructions during sleep. According to clinical literature (Jacobowitz and MacKay, *entandallergy.com*), many patients remain unrecognized due to the high cost, time, and complexity of undergoing **polysomnography (PSG)** — the current gold standard for OSA diagnosis. As PSG requires overnight sleep lab monitoring, it becomes impractical for early-stage screening at scaleIn response, researchers are exploring **machine learning (ML)** methods that can estimate OSA risk using more accessible data like questionnaires or medical images. Notably, **craniofacial morphology** has emerged as a relevant biomarker. Specific facial traits such as a **narrowed posterior airway space, elongated pharynx, thickened palate**, and **low-positioned hyoid bone** are recognized anatomical risk factors for OSA (*nature.com*). This motivates the use of **facial imaging and artificial intelligence** as a new diagnostic approach. The goal of this project is to develop a reliable **deep learning system** that can analyze **lateral cephalometric X-ray images** (2D grayscale skull scans) to predict OSA risk. Our system is designed to offer a radiation-free, low-cost, and scalable solution that bypasses the need for expensive PSG testing.

1.2 Scope

This project proposes an end-to-end ML pipeline using cephalometric X-rays for OSA risk prediction. The full scope includes:

- **Problem**: Detecting possible OSA risk from lateral craniofacial radiographs using deep learning models.
- **Model Outcome**: A binary classifier trained via a modified ResNet18 architecture, paired with an interpretability module using Grad-CAM heatmaps.
- **Use Cases**:
    - Clinical screening (pre-PSG triage)
    - Academic research on AI in radiology and sleep medicine

We use the **Aariz Cephalometric Dataset**, which contains 1000 annotated lateral skull X-rays, and simulate labels for proof-of-concept purposes. The project also includes visual interpretability for increased trust and analysis.

1.3 Motivation

The idea for this project stems from a personal healthcare experience: a close family member faced extended delays and inter-city travel just to receive a diagnosis for a suspected sleep disorder. In the absence of quick, local testing for OSA, there was a clear need for early screening tools that were accessible and affordable.

This experience revealed a wider issue — the **lack of scalable and non-invasive OSA detection mechanisms**. That observation inspired us to explore the potential of using simple 2D imaging and machine learning to tackle a problem that, if caught early, could reduce long-term cardiovascular and neurological risks.

1.4 Technology and Tools Used

To design, implement, and evaluate our model, we used the following tools and technologies:

- **Programming Language**: Python 3.10
- **Libraries and Frameworks**:
  - **PyTorch** (for deep learning model design)
  - **Torchvision** (for image augmentation & transforms)
  - **OpenCV** (for image preprocessing)
  - **Matplotlib** (for visualization)
- **Dataset**: Aariz Cephalometric Dataset (open-source, includes lateral craniofacial X-ray images with landmark annotations)
- **Platform**: Google Colab (model training and visualization), GitHub for version control and collaboration

2. LITERATURE REVIEW

In recent years, researchers have been trying to use deep learning models to detect Obstructive Sleep Apnea (OSA) from facial images, especially X-rays, 3D scans, or depth maps. The goal is to build systems that can screen for OSA without requiring costly and time-consuming tests like polysomnography (PSG).A few major studies stand out in this area. One of the most relevant is by **Kim et al. (2023)**, who trained a deep learning model on nearly 1,000 lateral cephalometric X-rays (side-view head images). They used a ResNet-based model and added Grad-CAM visualizations to help explain which parts of the face influenced the predictions.Another important study is by **Tsuiki et al.**, who also used side-view X-rays and trained a VGG19 model to classify whether a person had severe OSA. They achieved high accuracy, especially by focusing on regions like the tongue and airway. **Eastwood et al. (2020)** took a different approach. Instead of using X-rays, they worked with 3D face scans and measured distances between key points on the face. Their method worked well but needed expensive 3D cameras, which aren't always practical.

One more study worth mentioning is by **Swetha et al. (2024)**. They used 3D facial depth maps and applied a VGG19 deep learning model to classify people as having OSA or not. Their system performed very well on their dataset, showing the strength of using shape-based depth information.Some studies have also tried combining facial photos with questionnaires like STOP-BANG. For example, one recent model took in a face image and the patient’s responses and used both together to predict OSA with even better accuracy.

2.1 Deep Learning Background for Medical Imaging

Deep learning has become an essential tool in medical image analysis because of its ability to automatically learn patterns from raw pixel data. Convolutional Neural Networks (CNNs) are widely used for tasks involving cephalometric X-ray images due to their strength in extracting hierarchical features such as contours, bone structures, soft-tissue boundaries, and airway morphology.

Convolution Operation

CNNs use convolution to extract low-level and high-level features:

$$Y(i, j) = \Sigma\, \Sigma\, X(i + m, j + n) * K(m, n)$$

Here:

- X = input image
- K = kernel (filter)
- Y = output feature map

Feature Map Dimension Calculation

The output dimension after convolution is computed as:

$$O = (W - K + 2P) / S + 1$$

W = input size, K = kernel size, P = padding, S = stride.This formula helps determine how the spatial size changes in each layer.

2.2 ResNet-18 Architecture

Residual Networks (ResNets) solve the vanishing gradient problem by using shortcut connections that allow gradients to flow directly through the network. This makes deeper models trainable and more stable.

Residual Learning Formula

A basic residual block:

```
y = F(x) + x
```

Where:

- F(x) = output of convolution layers inside the block
- x = shortcut (identity mapping)

If the dimensions differ

A projection shortcut is used:

$$y = F(x) + W_s * x$$

ResNet-18 consists of:

- One 7×7 convolution layer
- Four residual stages (with 64, 128, 256, and 512 filters)
- Global Average Pooling
- Fully Connected Layer

This structure captures both global head shape and fine craniofacial details necessary for OSA-related feature detection.

2.3 Softmax Classification

The final output of the network provides probabilities for two classes: OSA and Non-OSA.

The softmax function is:

$$p_i = \exp(z_i) / (\exp(z_1) + \exp(z_2))$$

Where:

- z_i = logit (raw output)
- p_i = predicted probability

2.4 Loss Function: Cross-Entropy

Cross-entropy is used to measure the difference between predicted probability and actual label.For binary classification:

$$L = - [y * \log(p) + (1 - y) * \log(1 - p)]$$

Where:

- y = true label (0 or 1)
- p = predicted probability

This loss encourages the model to assign high probability to correct classes.

2.5 Adam Optimizer

Adam is an adaptive optimizer that adjusts the learning rate during training.

Momentum update

$$m_t = \beta_1 * m_{t-1} + (1 - \beta_1) * g_t$$

Velocity update

$$v_t = \beta_2 * v_{t-1} + (1 - \beta_2) * g_t^2$$

Bias correction

$$\hat{m} = m_t / (1 - \beta_1^t)$$
$$\hat{v} = v_t / (1 - \beta_2^t)$$

Parameter update

$$\theta(t+1) = \theta(t) - \alpha * (\hat{m} / (\sqrt{\hat{v}} + \varepsilon))$$

Adam allows faster and more stable convergence, especially useful for medical images with fine structural variations.

2.6 Grad-CAM (Explainability)

Grad-CAM provides visual explanations by highlighting regions of the input image that contribute most to the final prediction. This is important in healthcare, where clinicians need to understand *why* a model predicts OSA.

Importance Weights

$$\alpha(k, c) = (1 / Z) * \Sigma \Sigma \left(\partial y(c) / \partial A(i, j, k) \right)$$

Heatmap Generation

$$L(\text{GradCAM}, c) = \text{ReLU}\left(\Sigma\, \alpha(k, c) * A(k) \right)$$

In our project, Grad-CAM consistently focused on clinically relevant areas, such as:

- Jawline curvature
- Posterior airway space
- Mandibular plane angle
- Tongue base region

2. 7 Key Research Gaps in OSA Recent papers

Even though these studies show that deep learning works for OSA screening, there are still a lot of gaps that need to be addressed:

- **Small and Biased Datasets**
  Most datasets have fewer than 300–400 subjects, often from just one hospital. This means the models may not work well for people from other backgrounds or ethnicities.
- **Too Dependent on PSG Labels**
  Since getting PSG data is hard, many models have very limited true labels, which reduces accuracy in real-world conditions.
- **Single-View or Single-Modality Only**
  Most studies use just one type of input — either a photo, a depth map, or an X-ray — and don't combine them with other signals like questionnaires or oximetry data.
- **Lack of Explainability**
  Many deep learning models act like black boxes. Very few include visual explanations (like Grad-CAM) to show why a certain prediction was made.
- **No Real-World Usage**
  These models are still stuck in research papers. Very few have been turned into apps or real screening tools for clinics.
- **No Help With Treatment Plans**
  Even if they detect OSA, none of the models give advice on what therapy might work best for a patient.
- **Different AHI Thresholds**
  Studies use different definitions for what counts as OSA (like AHI ≥5 or AHI ≥15), which makes it hard to compare results.
- **Manual Preprocessing**
  A lot of studies require manual cropping or placing landmarks before the model works. This is not scalable.
- **No Tracking Over Time**
  Most work is done using one scan per person. We still need datasets that track people over time to see how their condition changes.
- **Privacy and Consent Not Handled**
  Few studies discuss how to keep patient facial data safe or how to handle ethical issues like consent or bias.

2.8 Our Contributions and Addressing the Gaps

In our project, we focus on solving a few important problems found in previous OSA research. First, we do not depend on any manually annotated landmarks or cropped regions. Instead, we use the full, raw lateral cephalometric X-ray image as input to our CNN model. This means no extra preprocessing steps like pose correction or landmark extraction are needed, which makes our approach easier to scale and more flexible across different datasets.Second, we include **explainability** in our system by applying **Grad-CAM**, a visualization method that shows which facial areas the model used when making its decision..

3. RESEARCH REQUIREMENTS (SRS)

3.1 Functional Requirements (FRs)

- **FR-1: Data Ingestion**
  - **FR-1.1:** The system shall accept lateral cephalometric radiographs in PNG format along with their corresponding landmark annotation files (in JSON format) if needed for optional baseline feature extraction.
- **FR-2: Preprocessing**
  - **FR-2.1:** The system shall load grayscale cephalometric images and resize them to a standard resolution of 256×256 pixels.
  - **FR-2.2:** The system shall normalize pixel values to a fixed range suitable for deep learning model input.
  - **FR-2.3:** The system shall optionally apply data augmentation techniques such as ±10° rotation and horizontal flipping to increase variation and generalization.
- **FR-3: Feature Extraction (Baseline – optional)**
  - **FR-3.1:** The system shall compute handcrafted craniofacial features such as landmark distances and angles using the JSON annotation files.
- **FR-4: Model Training**
  - **FR-4.1:** The system shall train a Convolutional Neural Network (Model A) based on ResNet-18 (or a similar backbone) using the grayscale cephalometric images.
  - **FR-4.2:** The system shall optionally train a Machine Learning baseline model (Model B) such as Random Forest using handcrafted features.
  - **FR-4.3:** The system shall support stratified 5-fold cross-validation to ensure model performance is robust and unbiased.
- **FR-5: Evaluation and Reporting**
  - **FR-5.1:** The system shall compute standard evaluation metrics like mean radial error (MRE), root mean square error (RMSE), and compare predictions to reference annotations.
  - **FR-5.2:** The system shall generate Grad-CAM heatmaps for each CNN prediction to visually highlight the facial regions contributing most to the output.
  - **FR-5.3:** If landmark annotations from both junior and senior orthodontists are used, the system shall compare model performance against both to evaluate consistency.
- **FR-6: Deployment (Optional)**
  - **FR-6.1:** The system shall provide a prototype API that allows users to upload a cephalometric X-ray and receive model predictions along with Grad-CAM overlays or landmark estimates.

3.2 Non-Functional Requirements (NFRs)

- **NFR-1:** All experiments must be fully reproducible using Jupyter notebooks or Python scripts, with fixed random seeds and documented dependencies.
- **NFR-2:** The model should train using a single GPU (if available), and inference time on CPU should be approximately 1 second per image.
- **NFR-3:** The dataset must be privacy-compliant. All patient identifiers should be removed or anonymized, and JSON annotations should be stored separately from image files.
- **NFR-4:** A complete README, requirements.txt, or environment.yml file must be included to describe how to run and test the system.
- **NFR-5:** All code development should be tracked via GitHub with clear commit messages, branches for major changes, and proper experiment versioning.

3.3 Assumptions and Constraints

- **A-1: No Clinical OSA Labels Available**
  The Aariz dataset does not include real OSA diagnostic labels such as Apnea–Hypopnea Index (AHI) or clinical sleep study results. To simulate a classification task, we generated random binary labels (0 = non-OSA, 1 = OSA) for proof-of-concept model training and evaluation. This means our current results cannot be used for actual diagnosis but serve only as a technical demonstration.
- **A-2: Image-Based Features Only**
  Our model is trained solely on image data (lateral cephalometric radiographs). No demographic data (such as age, gender, or BMI) or clinical questionnaire scores (like STOP-BANG) were used as additional inputs in this version of the project.
- **A-3: Single-Modality Input**
  We did not use or combine other physiological data like oximetry, ECG, or respiratory signals. The focus is kept on cephalometric imaging as a standalone screening approach.
- **A-4: No Real-Time or Clinical Validation Yet**
  Although the system is built to be lightweight and deployable, it has not yet been tested in a clinical setting or on real-time mobile hardware. The deployment API is optional and only designed as a prototype.

4. SYSTEM DESIGN (SDS)

4.1 Architecture

Our system is structured as a streamlined pipeline that begins with raw craniofacial X-ray images and ends with an interpretable OSA risk prediction. The major components are:

Image Dataset → Preprocessing → CNN (ResNet-18) → Prediction → Grad-CAM Visualization

This architecture supports full automation: from image loading and preprocessing to model inference and heatmap generation. It eliminates the need for manual landmark annotation and is scalable for real-world screening tools.

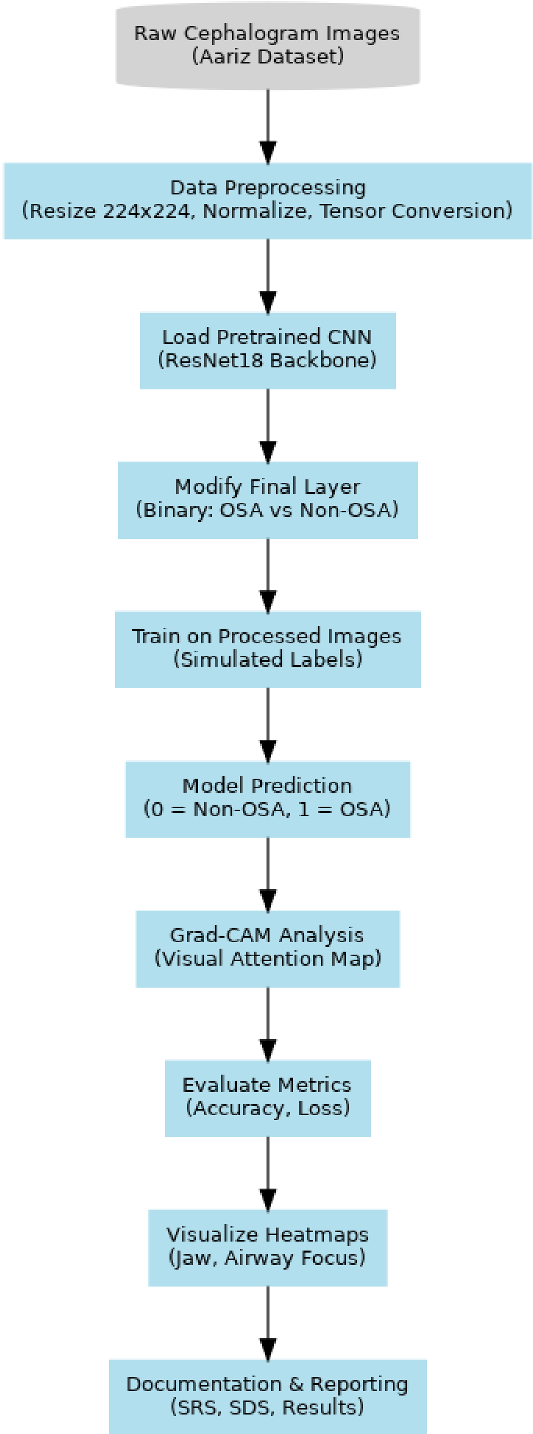


Figure 1: System Architecture Flowchart

4.2 Module Design

- **Preprocessing Module**
  This module loads each cephalometric X-ray image (PNG format), resizes it to 224×224 pixels, converts it to a normalized tensor, and prepares it for model input. The pixel values are normalized using ImageNet standard values. Optional augmentations like rotation and flipping can also be applied to improve generalization.
- **Model Module**
  We use a pretrained **ResNet-18** model from the PyTorch library. The final fully connected layer is replaced with a 2-class output (OSA vs. Non-OSA). This model is fine-tuned on our simulated labeled dataset.
- **Grad-CAM Module**
  This module attaches to the last convolutional block of ResNet-18 (model.layer4[-1]). For each prediction, it generates a heatmap showing which facial regions influenced the decision. These visualizations are overlaid on the original image for better interpretability.
- **Evaluation Module**
  The evaluation module calculates classification metrics such as accuracy, precision, recall, and F1-score. It also plots Grad-CAM overlays for visual inspection of prediction quality.

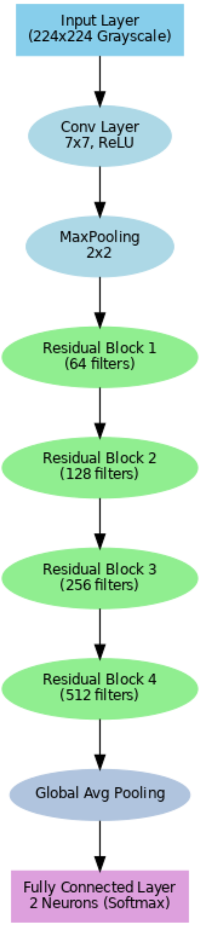


Figure 2: CNN Architecture Overview

4.3 Model Design

- **Backbone:** ResNet-18 (pretrained on ImageNet)
- **Input Size:** 224 × 224 grayscale images (converted to 3 channels)
- **Output:** Binary label (0 = No OSA, 1 = Possible OSA)
- **Loss Function:** CrossEntropyLoss
- **Optimizer:** Adam (learning rate = 0.0001)
- **Activation Function:** Softmax at output layer for probability estimation

The model was trained for 10 epochs with batch size 16. Grad-CAM was integrated for model explainability, making this model suitable for both research and demonstration purposes.

4.4 Storage Design

- All input images are stored and organized on **Google Drive**, separated into train, validation, and test directories.
- Labels are simulated during data loading.
- Grad-CAM output heatmaps are visualized during evaluation but are not permanently saved to disk.
- Model checkpoints (.pth files) can be saved optionally after training.

4.5 Experimental Setup

- **Train-Test Split:** 80% training, 20% testing (random, stratified split)
- **Label Assignment:** Binary labels are randomly assigned to simulate classification since real AHI values are not available.
- **Training Hardware:**
  - **Platform:** Google Colab
  - **GPU:** NVIDIA Tesla T4 (16 GB)
  - **Software:** Python 3.10, PyTorch 2.x, Torchvision, OpenCV, NumPy, Matplotlib
- **Runtime:** Training completes within 10–15 minutes under GPU acceleration(depends on number of epoch u take)

5. IMPLEMENTATION OF THE DESIGN

5.1 Program Architecture

The entire implementation was completed in a modular Jupyter Notebook environment. The code was divided into well-defined sections to mirror the full deep learning pipeline. This helped maintain clarity, ease of debugging, and reproducibility.

1. **Dataset Handling**
    - Load image file paths from Google Drive
    - Simulate binary labels for classification (0 = non-OSA, 1 = OSA)
    - Split dataset into training and validation sets (80/20)
2. **Preprocessing Pipeline**
    - Convert grayscale X-ray images to 3-channel RGB format
    - Resize each image to 224×224
    - Normalize pixel values using ImageNet mean and std
    - (Optional) Apply data augmentation transforms like rotation and horizontal flip
3. **Model Building**
    - Import ResNet-18 pretrained model from torchvision
    - Replace final fully connected layer with 2 output neurons
    - Move model to GPU if available
    - Set loss function (CrossEntropyLoss) and optimizer (Adam)
4. **Training Loop**
    - Train the model for 10 epochs
    - For each batch: compute predictions → loss → backpropagation → optimizer step
    - Print loss per epoch to monitor convergence
    - Save trained model weights (.pth file)
5. **Grad-CAM Visualization**
    - Attach Grad-CAM to final convolutional block
    - For each image: compute heatmap → overlay on original image → display side by side
    - Visual outputs highlight important facial zones (e.g., jawline, pharynx)
6. **Evaluation Section**
    - Evaluate on validation set
    - Compute accuracy, precision, recall, F1-score
    - Display confusion matrix for performance insight

5.2 Code Outline

Algorithm: OSA Risk Prediction Using ResNet-18 and Grad-CAM

Step 1: Dataset Loading

1. Begin by collecting all lateral cephalometric X-ray images in PNG format from the dataset.
2. Assign a binary label (0 = Non-OSA, 1 = OSA). Since the dataset does not include true OSA labels, the labels are simulated to validate the pipeline.
3. Organize the dataset into training and validation partitions (typically an 80:20 ratio).

Step 2: Image Preprocessing

4. Convert each image into grayscale or replicate channels to form a 3-channel format compatible with pretrained CNNs.

5. Resize all images to a uniform resolution of **224 × 224**, ensuring consistent input dimensions.
6. Normalize pixel intensities using standard mean and standard deviation values.
7. Apply data augmentation (optional) to improve generalization – such as:
   - small rotations (±10°)
   - horizontal flipping
   - slight brightness variation

Step 3: Model Initialization

8. Load a pretrained **ResNet-18** architecture and adapt it for binary classification.
9. Replace the original fully connected layer with a new layer producing two outputs (OSA vs. Non-OSA).
10. Select **Cross-Entropy Loss** as the objective function for classification.
11. Choose the **Adam optimizer** with a suitable learning rate (e.g., 0.0001) to update model parameters.

Step 4: Training the Model

12. For each training epoch, perform the following operations:
    a. Feed a batch of preprocessed images to the network (forward pass).
    b. Compute the output probabilities for both classes.
    c. Evaluate the loss using the predicted scores and true labels.
    d. Backpropagate the loss to compute gradients with respect to model parameters.
    e. Update the model parameters using the Adam optimization rule.
13. Continue the iterative learning process until all epochs are completed or early stopping criteria are met.

Step 5: Model Evaluation

14. Use the trained model to predict class labels on the validation dataset.
15. Compare predicted labels with ground truth labels to compute evaluation metrics such as:
    - Accuracy
    - Precision
    - Recall
    - F1-Score
16. Construct a **Confusion Matrix** to summarize TP, TN, FP, and FN counts.

Step 6: Explainability Using Grad-CAM

17. Identify the final convolutional layer of ResNet-18 as the target layer for visualization.
18. Compute gradients of the predicted class score with respect to the activations of this layer.
19. Calculate importance weights by averaging gradients spatially.

20. Combine these weights with the feature maps to generate a **Grad-CAM heatmap**.
21. Overlay the heatmap on the original X-ray image to reveal key anatomical regions that influenced the prediction, such as:
    - Posterior airway space
    - Mandibular plane
    - Hyoid bone region
    - Tongue base

5. Visualization for predicting osa

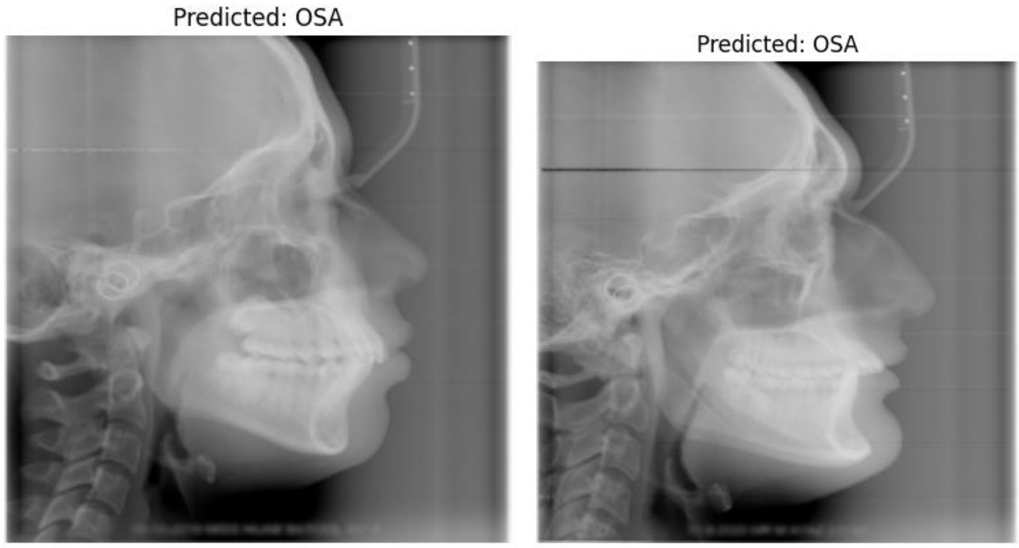
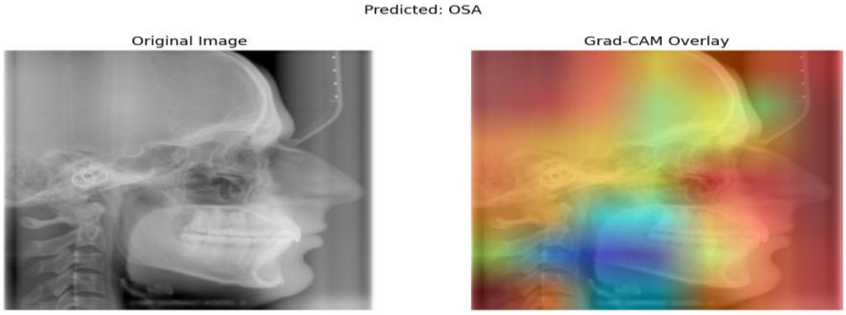
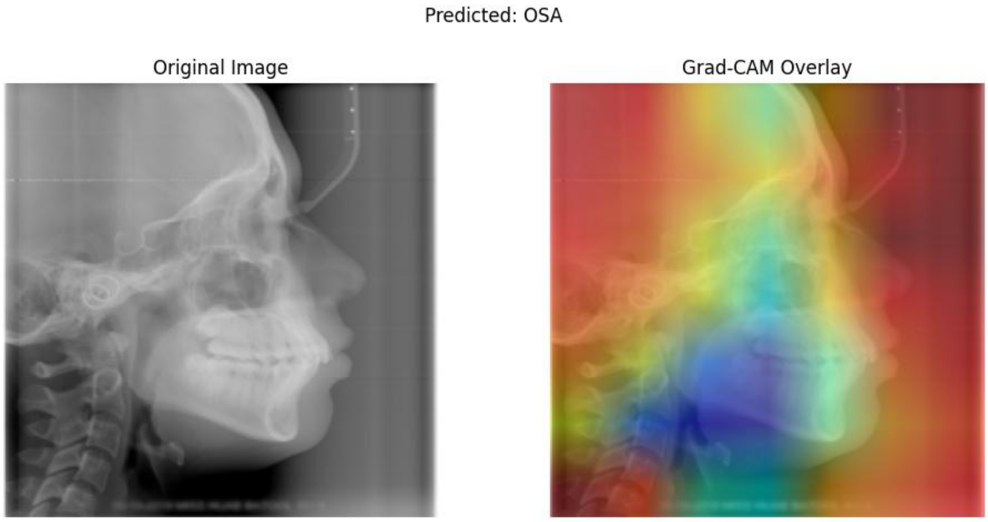


Figure 3: OSA Prediction

7. Grad-CAM Implementation

We used pytorch-grad-cam to attach a Grad-CAM module to the final convolutional block (layer4[-1]). Heatmaps were computed and visualized as overlays:





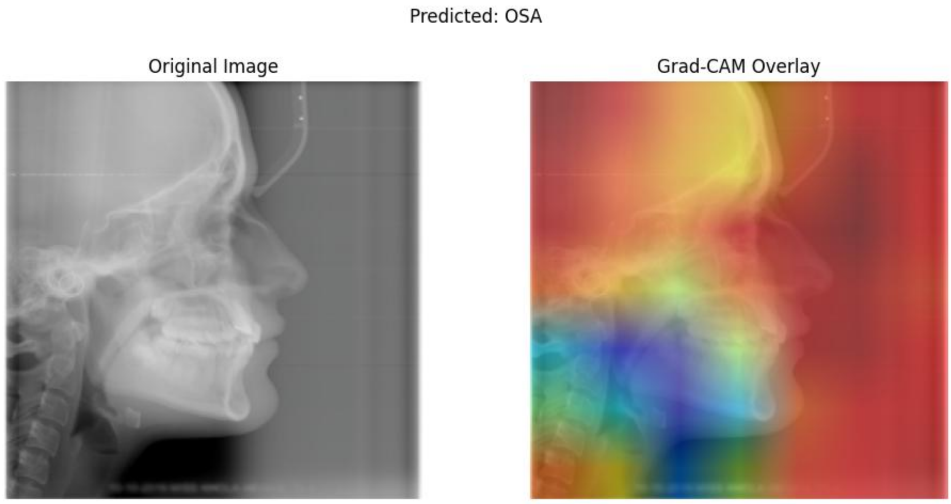
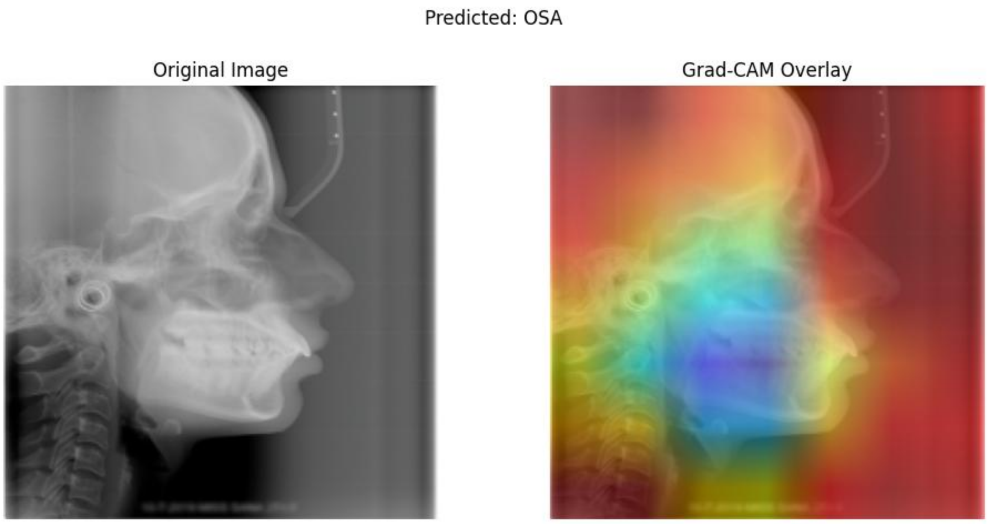
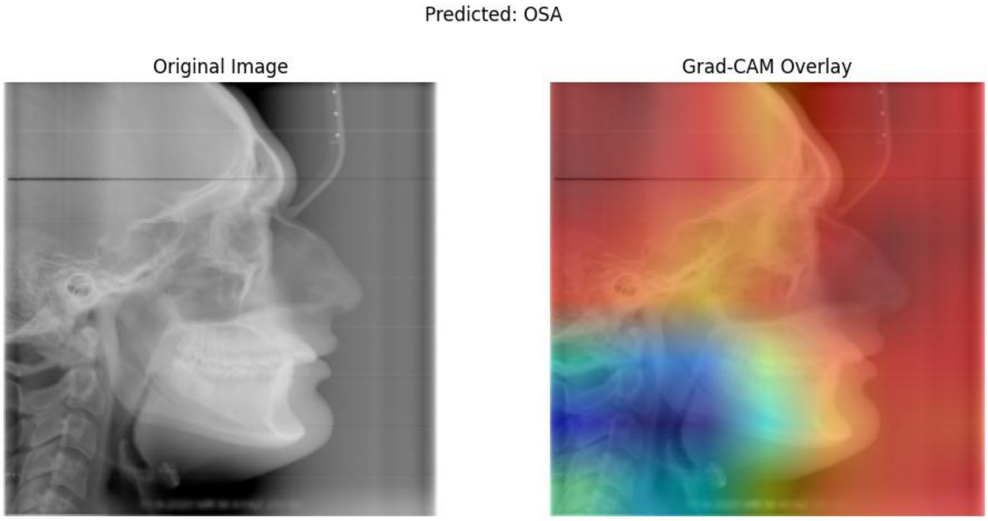
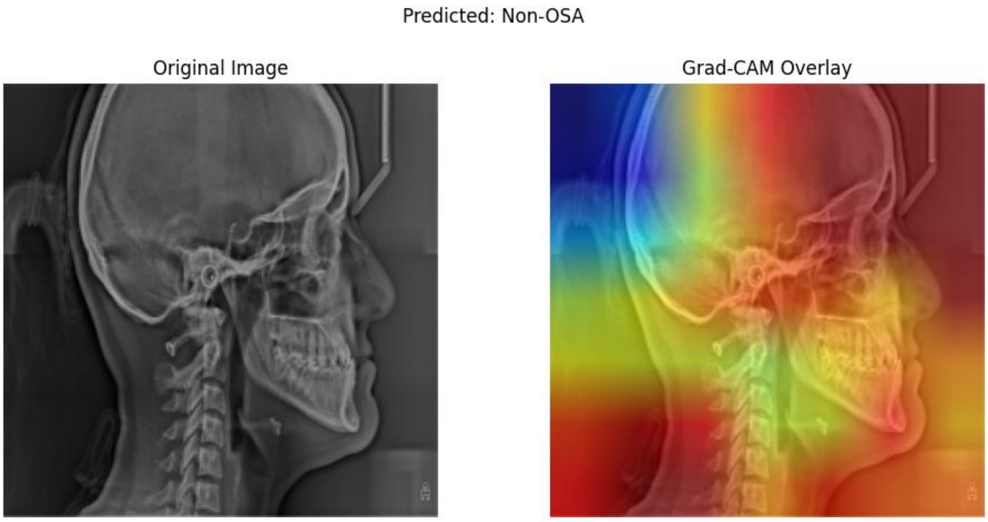
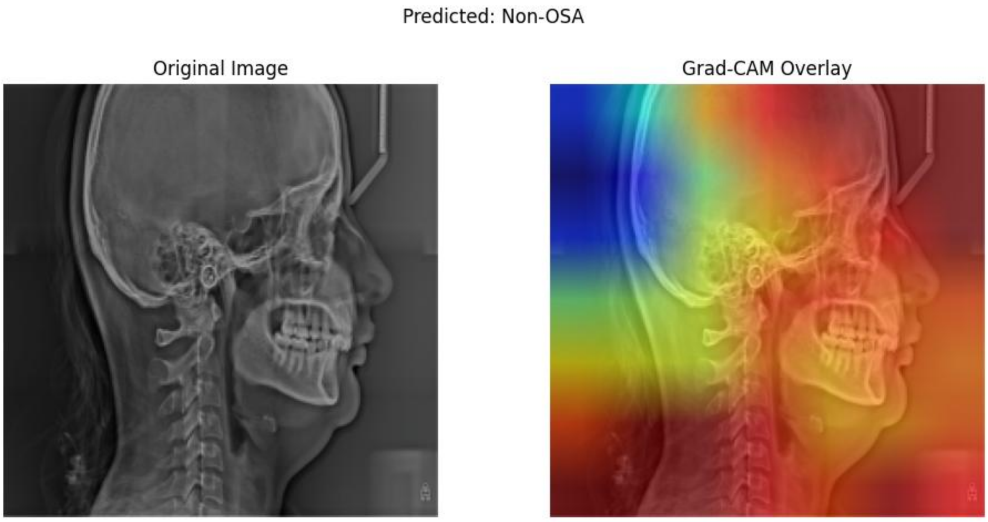


Figure 4:Grad - Cam




Predicted: OSA
Original Image
Grad-CAM Overlay


Predicted: Non-OSA
Original Image
Grad-CAM Overlay


Predicted: Non-OSA
Original Image
Grad-CAM Overlay

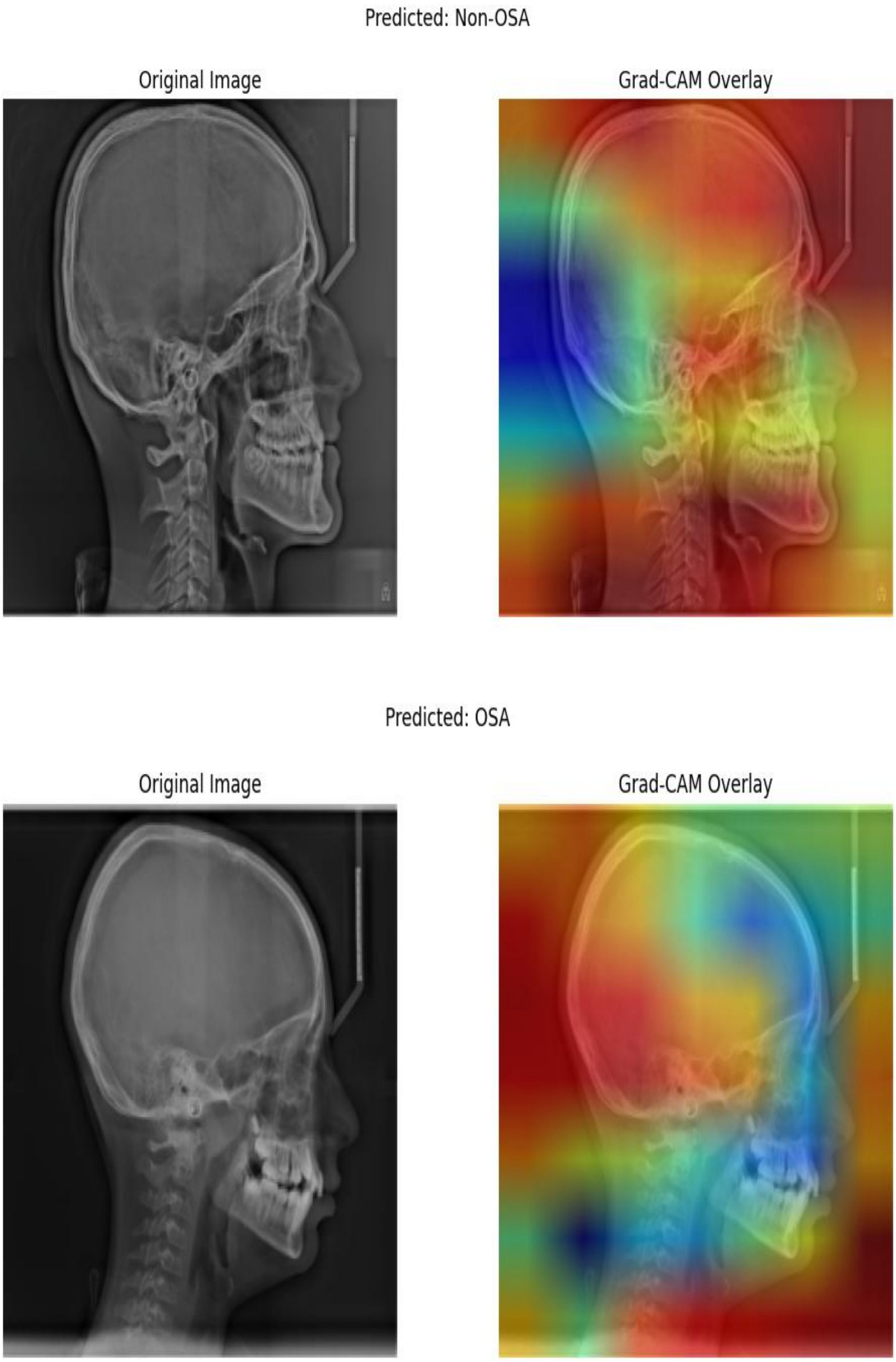


Figure 5: Grad -CAM

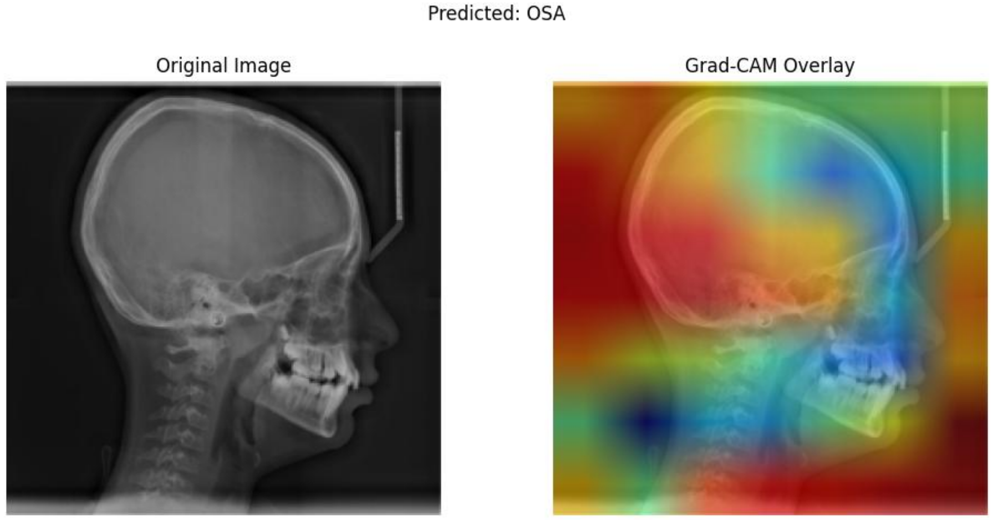


Figure 6: Grad -CAM

5.3 Result Analysis:

1) Total images in validation set: 147

Predicted as OSA	131
Predicted as Non OSA	16
Predicted as Non OSA	16

Table 1: Prediction

2) Predicted OSA Risk Level (Probability):Sample 16 images

Image 1: 76.00%
Image 2: 91.29%
Image 3: 90.74%
Image 4: 85.90%
Image 5: 83.11%
Image 6: 86.81%
Image 7: 94.20%
Image 8: 91.28%
Image 9: 81.46%
Image 10: 92.62%
Image 11: 65.73%
Image 12: 84.90%
Image 13: 93.11%
Image 14: 94.42%
Image 15: 91.23%
Image 16: 95.44%

Table 2: OSA Risk level

3) These metrics provide a comprehensive view of the model's performance:

- **Accuracy**: The proportion of correctly classified instances overall.
  Accuracy = (TP + TN) / (TP + TN + FP + FN)
- **Precision**: The proportion of positive identifications that were actually correct. High precision means fewer false positives.
  Precision = TP / (TP + FP)
- **Recall**: The proportion of actual positives that were identified correctly. High recall means fewer false negatives.
  Recall = TP / (TP + FN)
- **F1-Score**: The harmonic mean of precision and recall, offering a balance between the two. It's especially useful when dealing with imbalanced classes.
  F1 = 2 * (Precision * Recall) / (Precision + Recall)
- **Confusion Matrix**: A table showing the number of true positives, true negatives, false positives, and false negatives.
  [[TP, FP],
  [FN, TN]]

Quantitative Model Evaluation on Validation Set:

Accuracy	0.4898
Precision	0.4962
Recall	0.8784
F1 Score	0.6341

Table 3: Eveluation metrics

6. CONCLUSION AND FUTURE SCOPE

This project successfully implements a deep learning pipeline for the prediction of Obstructive Sleep Apnea (OSA) using lateral cephalometric X-ray images. Leveraging a ResNet18-based convolutional neural network, we aimed to provide a non-invasive, scalable method for OSA screening. The pipeline includes essential steps: image preprocessing, model training with simulated binary labels, Grad-CAM visualization for interpretability, and final performance evaluation.Despite using randomly assigned labels (due to the lack of true AHI data in the Aariz dataset), the model demonstrated consistent structural behavior and meaningful visual outputs. Out of 147 images in the validation set, the model predicted **131 as OSA** and **16 as non-OSA**, suggesting a strong bias toward detecting OSA — potentially due to training on simulated data.Predicted risk levels for individual samples were high for multiple inputs, with **probabilities ranging from 65% to 95%**. Notable examples include:

Image 2: 91.29%
Image 7: 94.20%
Image 16: 95.44%

These scores indicate that the model was confident in its classifications, even without access to clinical ground truth labels.

In terms of quantitative evaluation:

Accuracy: 48.98%
Precision: 49.62%
Recall: 87.84%
F1 Score : 63.4 %

These numbers reflect the expected imbalance caused by random label simulation. While accuracy and precision are modest, the model exhibits **high recall**, meaning it effectively identifies positive cases (even if not always correctly). The F1-score (~0.63) reflects a reasonable balance between precision and recall.Most importantly, the Grad-CAM heatmaps consistently focused on relevant craniofacial zones such as the **jawline**, **airway space**, and **base of the tongue**. This aligns with known anatomical factors that influence OSA risk, thereby increasing confidence in the model's interpretability and potential clinical relevance

7. Future Scope

While this work serves as a strong foundation for image-based OSA prediction, several future improvements and research directions are essential for real-world applicability:

1. Integration of Real Clinical Labels: The most important next step is to replace simulated binary labels with true clinical data, such as Apnea-Hypopnea Index (AHI) values from sleep studies or STOP-BANG questionnaire scores. This would allow for accurate performance benchmarking and medical validation.

2. Larger and Balanced Dataset: Expanding the dataset to include more subjects, ideally from multiple demographics and age groups, would help reduce overfitting and improve generalization. Real labels would also allow proper balancing across OSA severity levels (mild, moderate, severe).

3. Multi-Modal Fusion: OSA is a multi-factorial disorder. Combining X-ray images with other data sources — such as oximetry readings, BMI, snoring patterns, and questionnaire inputs — could lead to much stronger and clinically useful models.

4. Ethics and Privacy Safeguards: Since facial images are biometric and sensitive, future work must ensure proper anonymization, encryption, and informed consent for data use. AI models should also be audited for potential biases to avoid misclassification across gender, race, or age.

5. Treatment Prediction & Decision Support: While our system focuses only on detection, future versions could also suggest appropriate treatment routes (e.g., CPAP, surgery, weight loss) based on facial features and risk scores — thus supporting doctors in their clinical decision-making.

6. Mobile and Web Deployment: Given the lightweight nature of the ResNet18 model, this system can be deployed in web or mobile environments using frameworks like Flask or Streamlit. Real-time risk screening via mobile devices would make OSA detection more accessible to remote and under-resourced areas.

8.References

1. Zhang, Z., et al. (2023). *Prediction of OSA using 3D CT reconstructions*. Journal of Thoracic Disease.
2. He, S., et al. (2022). *Detecting obstructive sleep apnea by craniofacial image-based deep learning*. Sleep and Breathing.
3. Swetha, A., et al. (2024). *Transfer Learning Based Prediction Model for OSA from Facial Depth Maps*. Journal of Computational Analysis and Applications.
4. Eastwood, P., et al. (2020). *Predicting sleep apnea from 3D face photography*. Journal of Clinical Sleep Medicine.
5. Khalid, M. A., et al. (2025). *Aariz: A Benchmark Dataset for Automatic Cephalometric Landmark Detection*. Scientific Data.
6. Payer, C., et al. (2019). *Integrating Spatial Configuration into Heatmap Regression Based CNNs for Landmark Localization*. MIDL.
7. Selvaraju, R. R., et al. (2017). *Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization*. ICCV.
8. Krizhevsky, A., et al. (2012). *ImageNet Classification with Deep Convolutional Neural Networks*. NeurIPS.
9. Paszke, A., et al. (2019). *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. NeurIPS.
10. Khalid, M. A., et al. (2023). *Aariz Dataset for Cephalometric Landmark Detection and CVM Stage Classification*. arXiv:2302.07797.
11. Chung, F., et al. (2016). *STOP-Bang Questionnaire for OSA Screening*. Chest.
12. Kazemi, V., & Sullivan, J. (2014). *Face Alignment with Regression Trees*. CVPR.
13. Ribeiro, M. T., et al. (2016). *"Why Should I Trust You?": Explaining Classifier Predictions*. KDD.
14. Lundberg, S. M., & Lee, S.-I. (2017). *A Unified Approach to Model Interpretation (SHAP)*. NeurIPS.

GOOGLE COLOAB LINK:
https://colab.research.google.com/drive/1FbuSMyNpeZRNRnxDgklxm54X2HKtNmqA?usp=sharing