Project Report

Deep Learning Model for Classifying Snoring and Non-Snoring Sounds

CMSE 492: Applied Machine Learning

November 27, 2024

Aditya Pendyala

Table of Contents

Background and Motivation

Snoring is a common phenomenon that affects millions of people worldwide, often dismissed as a harmless annoyance. However, persistent snoring can indicate underlying health issues such as obstructive sleep apnea (OSA), a condition linked to increased risks of cardiovascular diseases, diabetes, and fatigue-related accidents. Identifying snoring patterns early and accurately is critical for timely intervention and treatment, improving individuals' quality of life and preventing severe health outcomes.

In medical contexts, accurate snoring detection systems play a vital role in diagnosing sleep disorders. Current solutions, such as polysomnography, are expensive, labor-intensive, and often require patients to stay overnight in sleep labs, leading to inconvenience and limited accessibility. A robust automated system capable of detecting snoring using audio data could transform the diagnostic process, making it faster, more affordable, and accessible to a wider population.

In everyday life, snoring detection has applications in consumer health devices, such as smartwatches and sleep trackers, to help users monitor their sleep quality. Integration of snoring detection into such devices can empower individuals to take proactive measures, seek medical advice when necessary, or adjust lifestyle habits to improve sleep.

This project aims to leverage machine learning techniques to develop an automated snoring detection system based on audio data. By analyzing audio representations—such as mel-spectrograms—I aim to identify the most effective features and models for snoring detection. The goal is to bridge the gap between advanced technology and practical applications, offering solutions that can benefit both clinical settings and personal health monitoring.

Objective

The objective of this project is to develop an efficient snoring detection system using transfer learning with the VGG16 model. Instead of building a model from scratch, we leverage the pre-trained VGG16 architecture to classify audio as either snoring or non-snoring.

The approach involves:

- Processing .wav audio files to extract key features such as mel-spectrograms.
- Visualizing these features as image plots, which serve as input for training the VGG16 model.
- Optimizing the model to enhance classification accuracy while reducing training time by reusing VGG16's learned features.

Metrics

In this project, I will evaluate the performance of my snoring detection model using precision, recall, and the F1 score, as these metrics are particularly well-suited for binary classification problems. Additionally, I will analyze the area under the precision-recall curve (AUC-PR) to gain a more comprehensive understanding of the model's ability to distinguish between snoring and non-snoring audio signals across different thresholds.

1. **Precision**:

   Precision measures the proportion of true positive predictions (correctly identified snoring instances) out of all positive predictions (both true positives and false positives). Precision is crucial in this context because false positives (non-snoring events misclassified as snoring) can lead to unnecessary interventions, false alarms in consumer devices, or misdiagnosis in clinical settings.

2. **Recall**:

   Recall (or sensitivity) measures the proportion of true positive predictions out of all actual positives (both true positives and false negatives). Recall is equally important because false negatives (missed snoring events) can result in undetected sleep disorders, delaying critical diagnosis and treatment.

3. **F1 score:**
   The F1 score is the harmonic mean of precision and recall, providing a single metric to balance both aspects. It is particularly useful when the dataset is imbalanced (e.g., more non-snoring samples than snoring samples). A high F1 score indicates a good trade-off between precision and recall, reflecting the overall effectiveness of the model.
4. **Area Under the Precision-Recall Curve (AUC-PR)**:
   The AUC-PR summarizes the trade-off between precision and recall across all possible thresholds. This metric is particularly valuable in scenarios where there is a class imbalance, as it focuses on the model's ability to identify positive cases (snoring) without being biased by the majority class (non-snoring).

Initial and Exploratory Data Analysis (EDA)

The dataset used for this project is the Snoring Dataset, which is publicly available on Kaggle. This dataset comprises audio recordings of snoring and non-snoring sounds, intended to help train and evaluate models for snoring detection.

- **Structure**: The dataset contains labeled audio files grouped into two folders: snoring and non-snoring.
- **File Format**: The audio files are in .wav format.
- **Class Distribution**: The dataset is balanced; comprising of five hundred .wav files for each class – snoring and non-snoring.
- **Data State and Issues:**
  - Among the 500 snoring samples, 363 samples consist of snoring sounds of children, adult men and adult women without any background sound. The remaining 137 samples consist of snoring sounds having a background of non-snoring sounds.

  - The 500 non-snoring samples consist of background sounds that might be available near the snorer. Ten categories of non-snoring sounds are collected, and each category has 50 samples. The ten categories are baby crying, the clock ticking, the door opened and closed, total silence and the minor sound of the vibration motor of the gadget, toilet flashing, siren of emergency vehicle, rain and thunderstorm, streetcar sounds, people talking, and background television news.
  - Raw audio waveforms are high-dimensional and less informative for direct input into models. Transformations such as mel spectrograms are required to extract meaningful features.
- **Cleaning and Transformations:**
  - Normalization: Scale audio signals to a consistent range to enhance model performance.
  - Transform audio data into mel spectrograms. These representations provide lower-dimensional and more structured inputs for machine learning models.
- **Feasibility for Goal:**
  - This dataset is well-suited for achieving the goal of snoring detection. Labels can be extracted from file names, making it effective for supervised learning.
  - The audio recordings contain rich information, allowing the extraction of meaningful features such as frequency patterns and temporal variations that are crucial for distinguishing between the two classes.
  - Additionally, the dataset's characteristics closely align with real-world scenarios, making the results directly applicable to medical diagnostics and consumer health applications, such as sleep monitoring devices.

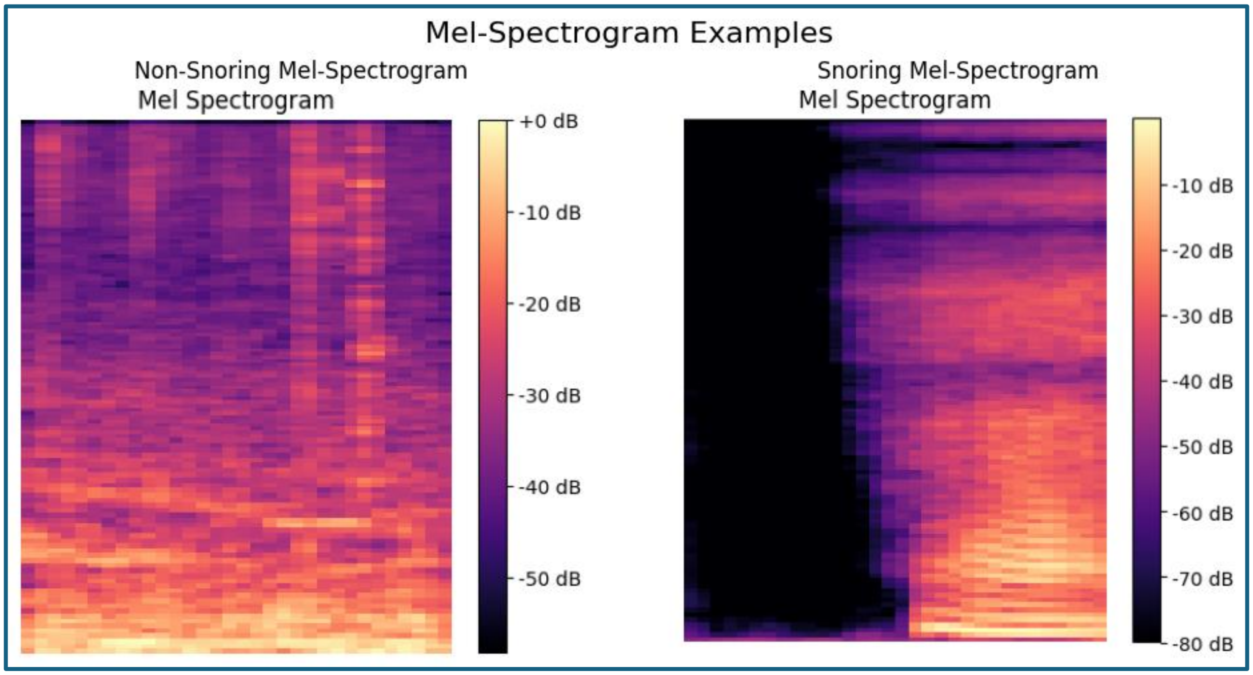

Figure 1: *Mel-spectrogram example plot for a non-snoring and snoring .wav file.*

Models

In this project, I am comparing two distinct machine learning models for snoring detection from audio data: **VGG16 with Transfer Learning** and **SGD Classifier**. The goal is to evaluate and compare the effectiveness of these models in classifying audio features, specifically Mel-spectrograms, into two categories: snoring and non-snoring.

- **VGG16 with Transfer Learning:**
  - **Architecture**: VGG16 is a deep Convolutional Neural Network (CNN) that has been pre-trained on a large dataset (ImageNet). We use this pre-trained model with transfer learning, which means we freeze the convolutional layers and add our custom fully connected layers on top. This allows the model to leverage the knowledge learned from a large image dataset and fine-tune it on our snoring detection task.

  - **Why it's used**: VGG16 is highly effective for image-based tasks and can capture spatial hierarchies in the input data. Since Mel-spectrograms are essentially time-frequency representations of audio, they can be treated as image-like data, making VGG16 an appropriate model to capture the relevant patterns for snoring detection.
  - **Advantages**: The model leverages transfer learning, which significantly reduces the time required for training and can lead to better performance, especially when there is limited data. VGG16 is known for its ability to extract complex patterns from images, which is beneficial for this task.

- **SGD Classifier:**
  - **Architecture**: The SGD (Stochastic Gradient Descent) Classifier is a linear model trained using stochastic gradient descent. In this case, we used it with the 'log_loss' loss function, which is suitable for binary classification tasks.
  - **Why it's used**: SGD is a simpler, more computationally efficient model compared to deep learning methods. It is a good baseline model to assess the performance of more complex models like deep neural networks. SGD is effective for large datasets and can work well when the decision boundary is relatively simple.
  - **Advantages**: The main advantage of SGD is its speed and simplicity. It can handle large datasets efficiently and is less prone to overfitting compared to more complex models, particularly when using regularization.

Training Methodology

Data Splitting:

- The dataset was divided into **training (70%)** and **testing (30%)** sets, ensuring balanced representation of both classes (snoring and non-snoring).
- Random stratified sampling was used to preserve the class distribution in the splits.

Model-Specific Training Details:

1. **SGD Classifier**:
    - **Feature Engineering**:
        - The Mel-spectrograms were flattened into feature vectors for input into the SGD model.
    - **Loss Function**:
        - **Logarithmic Loss** (log_loss) was used for binary classification.
    - **Regularization**:
        - **L2 regularization** was used to reduce overfitting.
    - **Cross-validation**:
        - 5-fold cross-validation was performed on the training set to evaluate and optimize hyperparameters (e.g., learning rate, alpha).
    - **Hyperparameter Tuning**:
        - A grid search was conducted over hyperparameters such as the regularization strength (alpha) and learning rate to find the optimal configuration.
2. **VGG16 with Transfer Learning**
    - **Preprocessing**:
        - Mel-spectrograms were generated from audio files, transforming audio data into image-like representations.
        - Images were resized to match the input shape required by VGG16 (224x224 pixels).
    - **Model Architecture**:
        - Utilized the pre-trained convolutional layers of VGG16 (trained on ImageNet), keeping them frozen to leverage learned feature representations.
        - Added custom fully connected layers, including a dense layer with 256 units and a dropout layer, for fine-tuning on the snoring classification task.
    - **Loss Function**:
        - Binary cross-entropy loss was used to optimize the model for binary classification.
    - **Optimizer**:
        - The Adam optimizer with default learning rate settings (0.001) was employed for efficient weight updates.
    - **Techniques to Avoid Overfitting**:

        - **Dropout Layer**: Incorporated a dropout layer with a 0.5 dropout rate to reduce overfitting.
        - **Early Stopping**:
            1. Training was halted when validation accuracy stopped improving for 5 consecutive epochs (patience = 5).
            2. Training would stop early if accuracy exceeded a predefined baseline of 95%, ensuring efficient use of resources.
    - **Learning Tracking**:
        - Loss and accuracy were plotted over training epochs for both training and validation data to monitor learning progress.
        - Validation metrics, including **Precision** and **Recall**, were tracked to ensure balanced performance across classes.
    - **Training Epochs**:
        - The model was trained for a maximum of 20 epochs, but early stopping was implemented to terminate training when validation performance plateaued.

Learning Curves and Performance Monitoring:

1. **VGG16 Learning Curves**:
    - Interpretation:
        - **Steady decrease in training and validation loss** indicates effective learning.
        - Validation loss plateaus while avoiding divergence, showing no significant overfitting.

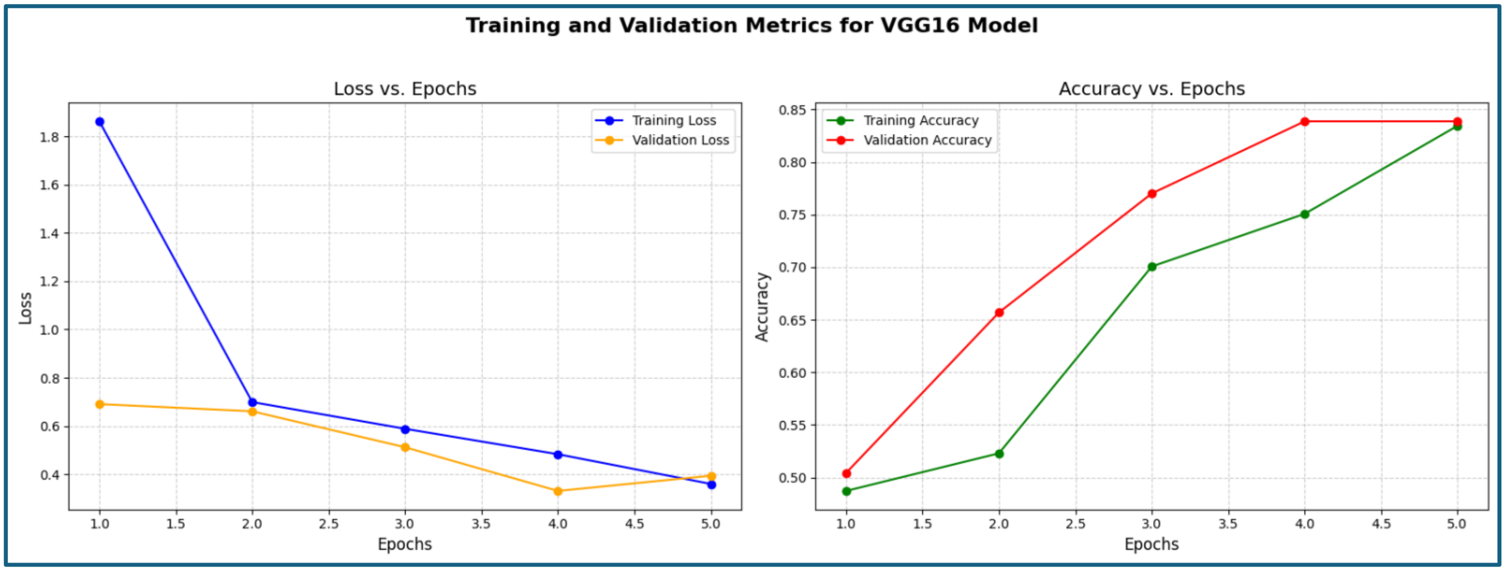

Figure 2: *Training and Validation Metrics for VGG16 Model*

2. **SGD Learning Curve**:
    - **ROC Curve**:
        - The ROC curve indicates modest classifier performance, with some ability to distinguish between classes.
        - The AUC score of **0.671** suggests the model performs better than random guessing but lacks strong predictive power.
        - Indicates potential issues such as suboptimal hyperparameters, insufficient features, or dataset imbalance.

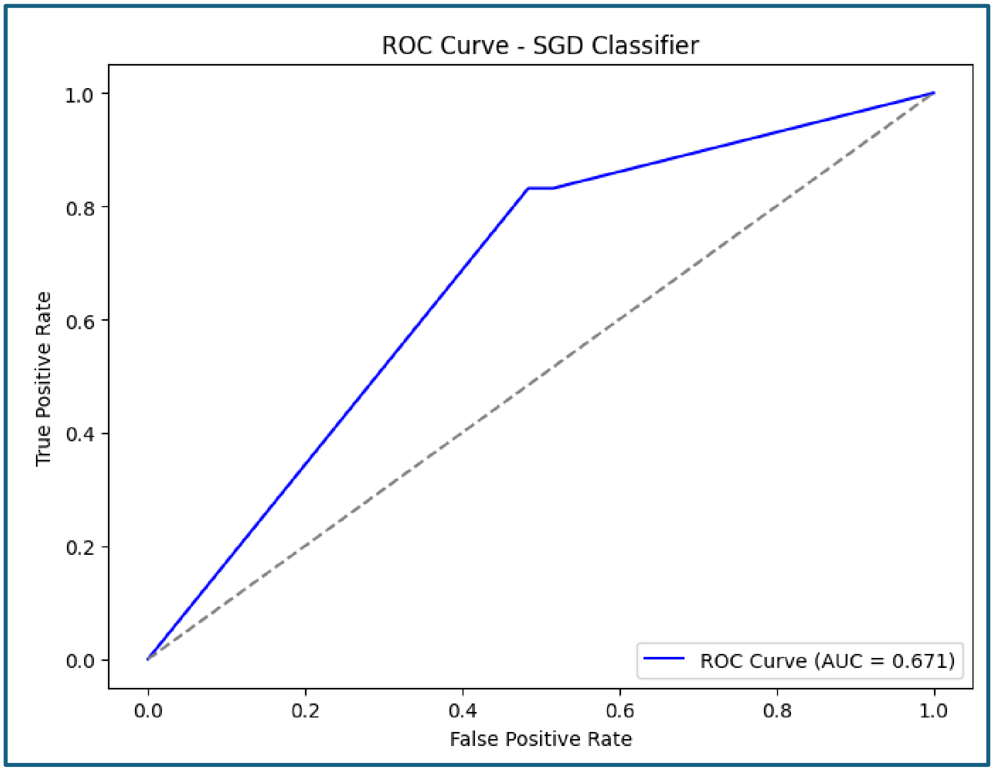

Figure 3: *ROC curve for SGD classifier.*

Results and Model Comparison

Performance Comparison:

- **VGG16**: The VGG16 model outperforms the SGD classifier in terms of AUC, precision, and recall. The AUC score of 0.98 indicates that the VGG16 model can effectively discriminate between snoring and non-snoring sounds with high confidence. The confusion matrix further confirms that VGG16 achieves a good balance between detecting snoring and non-snoring sounds, with relatively few false positives and false negatives.
- **SGD Classifier**: While the SGD classifier performs decently, its performance is less robust compared to the VGG16 model. The accuracy of 0.68 suggests that it struggles to identify snoring and non-snoring sounds with high precision. The recall for non-snoring (0.52) indicates that the SGD classifier is less reliable in detecting non-snoring sounds, which could be a significant drawback in real-world applications.

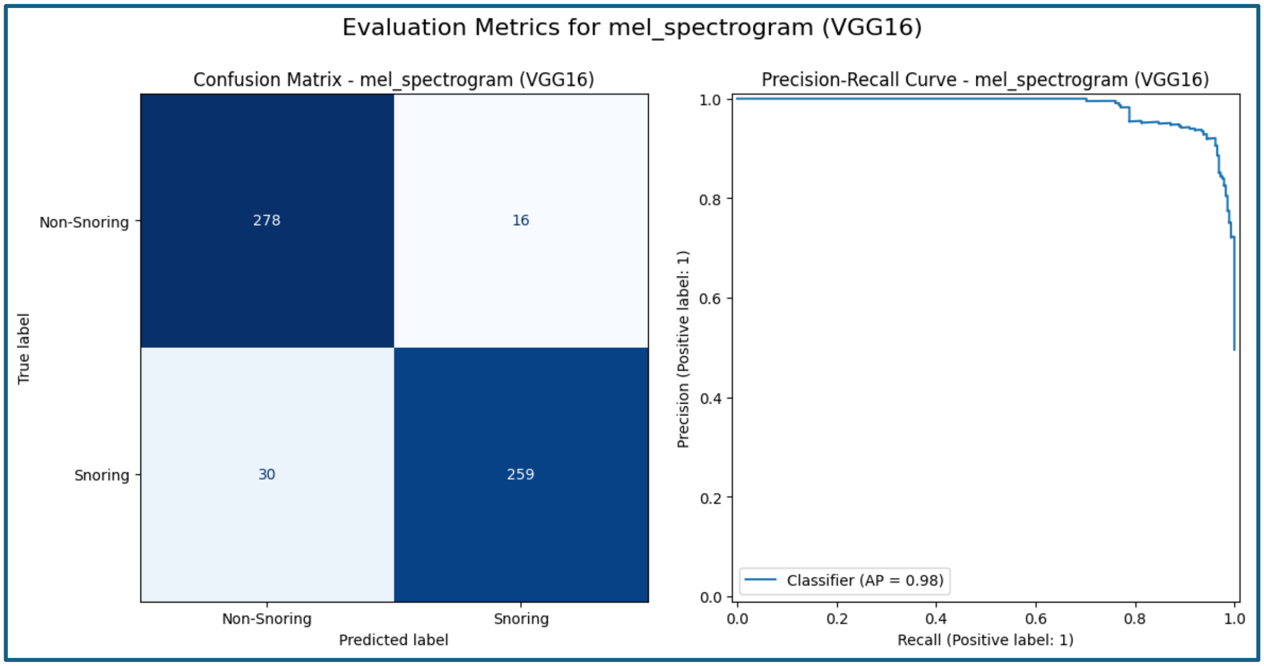

Figure 4: *Evaluation metrics for VGG16 model.*

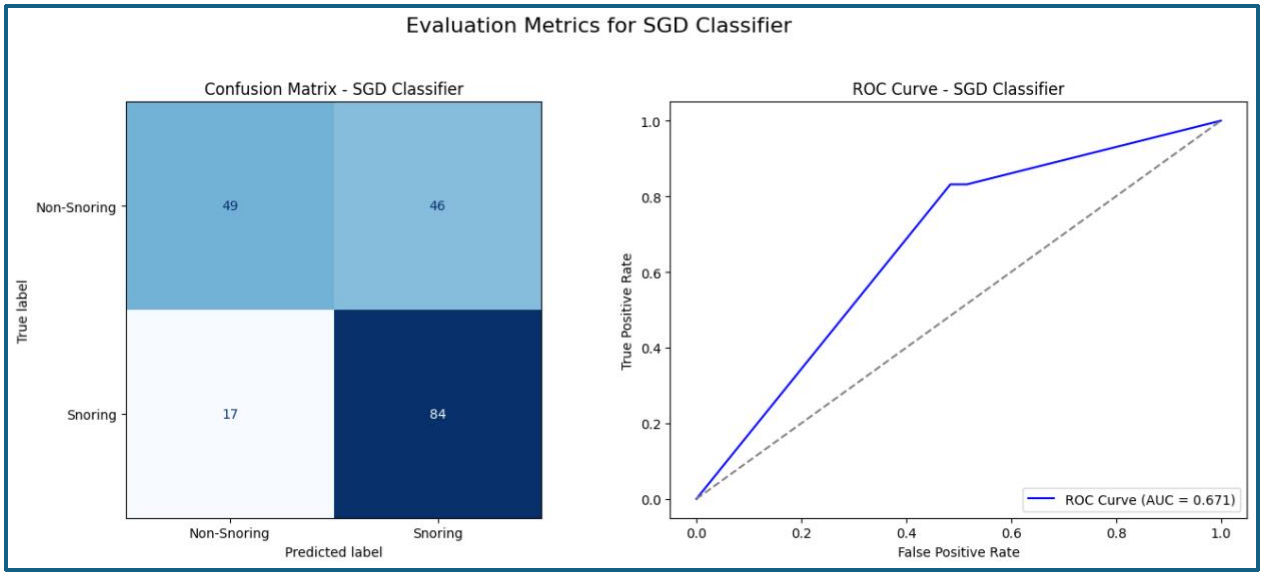

Figure 5: *Evaluation metrics for SGD classifier.*

Conclusion

The VGG16 model significantly outperforms the SGD classifier in this task. The high AUC and balanced confusion matrix demonstrate that the VGG16 model is highly effective for this type of audio classification task, likely due to its ability to learn complex patterns in the Mel-spectrograms. On the other hand, the SGD classifier, while useful as a baseline model, is less effective at handling the complexity of snoring detection, particularly for the non-snoring category. Thus, the VGG16 model is the better choice for this task, but the SGD classifier serves as a useful comparison to highlight the power and convergence of deep learning models in solving more complex classification problems.

Reference

T. H. Khan, "A deep learning model for snoring detection and vibration notification using a smart wearable gadget," Electronics, vol. 8, no. 9, article. 987, ISSN 2079-9292, 2019.