CHAPTER I
INTRODUCTION

1.1 Background

The ear constitutes a vital organ in the human body because it performs essential functions in auditory perception and contributes to the maintenance of body balance. The importance of this organ's role requires attention and proper care to maintain ear health. However, to date, public awareness of ear health remains relatively low. This situation has led to ear disorders becoming a fairly common health issue across various age groups. According to the WHO, in 2021, approximately 430 million individuals worldwide required rehabilitation services for hearing loss in 2021, of whom approximately 109.4 million were from Southeast Asia. The global burden is expected to increase substantially, with projections indicating that as many as 2.5 billion people could be affected by hearing loss by 2050 [1]. In Indonesia, this challenge is further intensified by inadequate public awareness regarding the importance of early detection and by the unequal availability of ear, nose, and throat (ENT) healthcare services, particularly across rural and remote regions. These conditions lead to delayed diagnosis, which can potentially worsen hearing problems and increase the risk of permanent hearing loss [2].

Ear disorders encompass a variety of conditions that require different treatments, such as earwax plugs (cerumen impaction), myringosclerosis, and chronic otitis media (COM). Each of these disorders exhibits specific visual characteristics when observed through an otoscope, so this information can be utilized for image-processing-based classification [3]. Although each ear disorder has distinct visual characteristics, the diagnostic process still requires a high degree of precision because these differences are not always easy to identify in clinical practice. Manual otoscopic examination is subjective and highly dependent on the examiner's experience, so its accuracy is often inconsistent [4]. A study reported that the accuracy rate for diagnosing ear infections by general practitioners ranges from 50–70%, while that of ENT specialists is approximately 72–82%. This difference indicates variability in the interpretation of otoscopic images, which is influenced by clinical experience and the conditions under which the examination is conducted [5]. This variability has the potential to lead to differences in interpretation during the diagnostic process, particularly in primary care facilities that have limited diagnostic equipment and a shortage of specialists. Consequently, artificial intelligence-based otoscopic image classification technology has emerged as an important approach for supporting the diagnostic process, with the potential to enhance the consistency and efficiency of clinical examinations. Beyond improving the consistency of examination outcomes, this automated approach has the potential to facilitate early screening, minimize diagnostic errors, and increase access to healthcare services, particularly in regions with limited medical resources [6].

Advances in medical image processing technology and artificial intelligence have created new opportunities to overcome the challenges of manual diagnosis. Convolutional Neural Networks (CNNs), a deep learning architecture widely applied in computer vision, have shown strong capability in analyzing medical images with high accuracy [7]. Their ability to automatically learn and extract discriminative visual features from image data has made CNNs one of the most extensively adopted deep learning models for medical image analysis [8].

As the need for more efficient models has grown, various modern CNN architectures have been developed, including EfficientNet-B0, which is designed to achieve optimal feature extraction with minimal computational complexity. This architecture is relevant to this research because it can extract important features from images more efficiently than other CNN models [5]. This study employed EfficientNet-B0 using a transfer learning approach, where a model pre-trained on a large-scale dataset was leveraged to identify visual patterns in medical images, including otoscopic images. The extracted features were then classified using the XGBoost algorithm.

Extreme Gradient Boosting (XGBoost) is an ensemble learning algorithm built upon the gradient boosting framework, has demonstrated strong performance in various machine learning applications, including medical image classification [9]. This algorithm offers several notable advantages, including the use of regularization techniques to reduce the risk of overfitting, support for parallel computation that improves training efficiency, and the capability to generate interpretable feature importance scores that can assist medical practitioners in understanding the model's predictions [10]. Integrating Convolutional Neural Networks (CNNs) as feature extractors with XGBoost as the classification algorithm has the potential to develop models that achieve high predictive accuracy while maintaining interpretability and computational efficiency [9].

Viscaino et al. (2020) developed a computer-aided diagnosis system to assist general practitioners in identifying ear disorders, particularly in healthcare settings where access to otolaryngology specialists is limited. The study evaluated the performance of three machine learning algorithms, namely Support Vector Machine (SVM), K-Nearest Neighbor (k-NN), and Decision Tree, in combination with three feature extraction techniques: Color Coherence Vector (CCV), Discrete Cosine Transform (DCT), and Filter Bank. The experimental dataset comprised images representing three ear conditions myringosclerosis, earwax plug, and chronic otitis media (COM) with 720 images allocated for training and validation and 160 images reserved for testing. Experimental results demonstrated that the SVM and k-NN models achieved higher classification accuracy than the Decision Tree model. Among the evaluated methods, the SVM classifier produce the best overall performance, achieving an average accuracy of 93.9%, sensitivity of 87.8%, specificity of 95.9%, and positive predictive value (PPV) of 87.7%, indicating its potential to support clinical decision-making in primary healthcare settings [3].

Bingol (2024) developed a hybrid deep learning framework designed to support physicians in diagnosing Otitis Media with Effusion (OME), a middle ear disorder that is often difficult to identify through conventional clinical examination. The study employed autoendoscopic images and extracted image features using two CNN architectures, namely EfficientNet-B0 and DenseNet201, using both the original images and those enhanced through the Gaussian method. The extracted features were subsequently fused and refined through Neighborhood Components Analysis (NCA) to reduce feature dimensionality while preserving the most informative representations. The optimized feature set was then evaluated using several machine learning classifiers, where Support Vector Machine (SVM) demonstrated the best performance, achieving an accuracy of 98.20%. These findings suggest that integrating CNN-based feature extraction with machine learning classification can provide highly effective performance for the detection of OME [11].

There have been studies in other medical fields that have performed classification by implementing a combination of CNN and XGBoost algorithms. Kumar et al. (2024) proposed a hybrid classification framework for brain tumor identification from MRI images by employing EfficientNet-B0 as the feature extraction model and XGBoost as the classification algorithm. This model leverages the parameter efficiency of EfficientNet-B0 and XGBoost's ability to produce stable and robust classification decisions. The model incorporated several preprocessing procedures, including resizing, normalization, and data augmentation, together with architectural techniques such as Global Average Pooling and dropout. This configuration resulted in outstanding performance on the Kaggle dataset, achieving an accuracy of 99.21%, precision of 99.18%, recall of 98.96%, and an F1-score of 98.92%. The study's findings demonstrate that the EfficientNet-B0–XGBoost hybrid technique can increase medical image categorization accuracy and efficiency [8].

Hedhoud et al. (2023) proposed a hybrid framework for pneumonia classification by integrating Convolutional Neural Network (CNN)-based feature extraction with XGBoost as the classification algorithm. The model was developed and evaluated using chest X-ray images obtained from the Guangzhou Women and Children's Medical Center to differentiate bacterial and viral pneumonia, two conditions that are often difficult to distinguish through clinical assessment and frequently require the expertise of radiologists for accurate diagnosis. Evaluation under multiple testing scenarios demonstrated that the proposed CNN–XGBoost model achieved competitive performance, obtaining an accuracy of 87%, specificity of 89%, and sensitivity of 85%. In addition, the model exhibited a rapid prediction time of approximately 7 seconds, indicating its computational efficiency. The study's findings suggest that integrating CNN and XGBoost can enhance automated pneumonia diagnosis' efficacy and function as a practical substitute when medical staff is scarce [12].

Based on this, this research attempts to create an automatic classification system for ear disorder images using EfficientNet-B0 as the feature extractor and XGBoost as the classifier for four classes: Normal, Earwax Plug, Myringosclerosis, and Chronic Otitis Media (COM). Variations in image quality, shooting angles, and light reflections often lead to inconsistent manual identification. To address this issue, the preprocessing stage incorporates Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance local image contrast, thereby improving the visibility of image details, particularly in low-contrast regions, before feature extraction is performed by EfficientNet-B0. Similar preprocessing strategies have demonstrated effective performance across various medical imaging applications. For example, in COVID-19 detection from chest X-ray images, Contrast Limited Adaptive Histogram Equalization (CLAHE) has been applied to enhance the visibility of image features before classification is performed using a hybrid CNN–XGBoost model [13]. Thus, the combination of CLAHE, EfficientNet-B0, and XGBoost is expected to produce more accurate, stable, and reliable classification results.

1.2 Research Questions

In light of the background presented above, the research questions addressed in this study are formulated as follows:

1. How can the combination of EfficientNet-B0 and XGBoost be applied to classify ear disorder images?
2. How does the hybrid EfficientNet-B0 and XGBoost model perform in terms of classification accuracy for ear disorder images?

1.3 Research Objectives

In accordance with the formulated research questions, this study aims to:

1. To implement the combination of EfficientNet-B0 and XGBoost for ear disorder image classification.
2. To assess the performance and classification accuracy of the hybrid EfficientNet-B0–XGBoost model in classifying ear disorder images.

1.4 Research Benefits

By adopting XGBoost as the classification model and EfficientNet-B0 as a feature extractor, this study is anticipated to contribute to the advancement of image-based ear disorder classification using otoscopic images. The combination of these methods is expected to improve classification accuracy and overall model performance while demonstrating the effectiveness of EfficientNet-B0 in extracting discriminative image features to support the classification process performed by XGBoost.

1.5 Scope and Limitations

To ensure that this study remains focused on the research objectives, the following limitations are defined:

1. he study utilizes a secondary dataset sourced from the Ear Imagery Database, which is publicly available through the Figshare platform https://figshare.com/articles/dataset/Ear_imagery_database/11886630?file=21793293.
2. The dataset consists of otoscopic ear images belonging to four classes: Normal, Earwax Plug, Myringosclerosis, and Chronic Otitis Media (COM), with a total of 880 images. These data were collected from a previous study.
3. The classification process is performed using RGB otoscopic images without considering patients' clinical data or medical records.
4. The dataset was generated by extracting frames from otoscopic videos; therefore, similar images may exist within the dataset. Since patient identities and original video information are unavailable, the dataset is split at the image level, which may introduce data leakage due to similarities among frames.
5. This study does not address the implementation of the proposed system in clinical practice or its integration with medical devices.