CHAPTER 1

INTRODUCTION

1.1 Background

Respiratory diseases represent a major cause of morbidity and mortality worldwide. According to World Health Organization (WHO) data, conditions such as Acute Respiratory Infections (ARI), pneumonia, asthma, tuberculosis, and Chronic Obstructive Pulmonary Disease (COPD) continue to account for high annual death tolls, particularly in developing countries.[1], [2] A report by *The Lancet* on Chronic Respiratory Diseases notes that conditions including asthma and COPD are among the leading causes of morbidity in the global adult population.[3] In Indonesia, ARI is one of the most frequent reasons for visits to primary healthcare facilities. This situation highlights the importance of early detection efforts as a preliminary step toward prevention, ensuring patients receive appropriate medical intervention promptly. Health surveys in Indonesia show that the incidence of Acute Respiratory Infection (ARI) reached 4.8% in 2022.[4] ARI (including acute upper respiratory infections and pneumonia) frequently manifests as an acute respiratory disorder and leads to high healthcare utilization.

The most common symptom in patients with respiratory diseases is coughing. Coughing is a natural defense mechanism of the body that serves to clear the respiratory tract of secretions, mucus, or foreign particles, making it an early indicator of respiratory system disorders. Cough symptoms often serve as an early indicator of respiratory disorders, ranging from mild to severe.[5] Cough conditions can be categorized into two main types: dry cough (non-productive) and wet cough (productive). The difference between these types of cough is linked to specific clinical conditions, where a wet cough is generally associated with the accumulation of secretions in the respiratory tract, such as in cases of pneumonia and lower respiratory tract infections. A dry cough is frequently associated with conditions such as asthma, allergies, or viral infections, whereas a wet cough is generally related to diseases such as acute respiratory infections (ARI), bronchitis, pneumonia, and tuberculosis, which are characterized by excessive sputum production. However, in clinical practice, the process of identifying the type of cough often remains subjective and dependent on medical staff observations as well as patient complaints, creating the potential for differences in interpretation. Therefore, distinguishing between a dry cough and a wet cough is an important aspect, as each type can provide early clues regarding the underlying pathological condition in the respiratory system and serve as the basis for determining subsequent medical management.[4], [5]

Advances in Artificial Intelligence (AI) technology, particularly Machine Learning and Deep Learning, have opened new opportunities in the medical field, especially for voice data processing. In the field of cough analysis, several studies have utilized cough sound recordings combined with Machine Learning methods to detect COVID-19 and have reported promising performance from cough-based models.[6] The Mel-Frequency Cepstral Coefficients (MFCC) audio feature extraction method is widely used because it can represent sound characteristics by mimicking the human auditory system. Across various audio signal processing studies, MFCC has proven effective in extracting informative spectral features for classification and pattern recognition tasks, including cough sounds, music, and human speech.[7] On the other hand, the Capsule Network (CapsNet) is a deep learning architecture capable of preserving spatial relationships between features through vector representations and dynamic routing mechanisms, making it generally more reliable at recognizing complex patterns compared to conventional CNNs that rely on pooling operations.[8] Furthermore, in the field of audio signal processing, MFCC is a popular feature extraction method due to its ability to mimic human perception of sound frequencies, rendering the resulting features more relevant to acoustic characteristics than standard linear transformations.[9] Many studies in sound classification utilize MFCC as a foundational feature because this representation is stable against signal variations and captures essential spectral components across diverse audio classification contexts. [7]

Although various studies have successfully used CNNs, RNNs, or LSTMs in cough sound classification, these approaches still have limitations, particularly in capturing complex relationships between spectral features in non-stationary audio data with high noise levels. CNN models tend to lose spatial relationships between features due to the pooling process, whereas RNNs focus more on temporal dependencies but are less effective at preserving the spatial structure of audio features. In fact, cough sounds possess interconnected spectral and temporal patterns that are essential for distinguishing the characteristics of dry and wet coughs.[10], [11], [12], [13] . Therefore, an architecture capable of preserving these inter-feature relationships more completely is required. Capsule Networks (CapsNet) offer advantages through their routing-by-agreement mechanism, which allows hierarchical relationships between features to be maintained, making them theoretically better suited to handle the complexity of cough sound patterns compared to conventional CNNs or RNNs. The study *"Improving Performance and Inference on Audio Classification Tasks Using Capsule Networks"* demonstrates that CapsNet can achieve better performance than traditional CNNs in audio classification. Furthermore, research combining CapsNet with audio features highlights CapsNet's potential for sound-related tasks, such as audio event detection and voice classification. [8]

In addition to the choice of classification model, audio feature representation plays a crucial role in determining the system's success. Mel-Frequency Cepstral Coefficients (MFCC) are one of the most widely used feature extraction methods in sound signal processing due to their ability to represent acoustic spectral characteristics aligned with human hearing; these coefficients are generated through a Mel-scale spectral transformation that mimics human frequency perception responses to audio signals.[14] MFCC has been extensively applied as a foundational feature in various audio classification studies, including music and other pattern recognition tasks, owing to its representational stability against frequency variations and noise.[15] On the other hand, Capsule Networks represent an advanced deep learning architecture capable of preserving spatial relationships between features through vector representations and dynamic routing mechanisms, which have been shown to be effective for audio classification tasks with complex feature hierarchies and offer advantages in maintaining part-to-whole relationships compared to standard CNN/RNN models.[8] In most previous studies, MFCC features were typically processed using CNNs or RNNs, which still present limitations in preserving relationships between spectral features. This study integrates MFCC with Capsule Networks to leverage MFCC's strengths in representing cough sound characteristics alongside CapsNet's capability to maintain complete inter-feature relationships. This combination is expected to produce a more stable and accurate classification process for dry and wet coughs, particularly on cough data exhibiting high variability and containing noise.

This study utilizes a single data source: secondary data obtained from the Kaggle public dataset *"Cough Sound Dataset for Covid-19 Detection."* This dataset contains cough sound recordings from various individuals, totaling 27,550 entries, accompanied by metadata such as cough condition labels, subject identity, and other supporting information used in the annotation process.[16], [17] Out of the overall dataset, 2,348 entries were labeled as dry cough (*dry*) or wet cough (*wet*), aligning with the scope of this research. These data were not used directly as experimental data; instead, they underwent label selection, label conflict checks, audio file verification, and audio quality audits. Based on the data curation process during project implementation, a final experimental dataset in the form of a balanced, confident subset of 800 audio samples was obtained, consisting of 400 dry cough samples and 400 wet cough samples. A balanced dataset was selected to minimize model bias toward the majority class while ensuring the objectivity of the training and classification evaluation processes. This final dataset was subsequently used in the training, validation, and testing data splitting process to construct and evaluate the classification model.[17] By utilizing this curated dataset, this research builds a cough classification model using a combination of Mel-Frequency Cepstral Coefficients (MFCC) and Capsule Networks. The resulting system is intended to serve as an early screening tool to differentiate between dry and wet coughs based on acoustic characteristics. The classification results are strictly for informational purposes, intended to provide initial information, and are not intended for medical diagnosis or as a substitute for examination by healthcare professionals.

1.2 Problem Statement

Based on the background described, the problem statements for this research are formulated as follows:

1. How can data selection, audio quality auditing, audio preprocessing, and Mel-Frequency Cepstral Coefficients (MFCC) feature extraction be performed to generate consistent cough sound data ready for the classification process?
2. How can an MFCC-based Capsule Network (CapsNet) model be designed and implemented to classify dry and wet coughs on a balanced, high-confidence dataset?
3. What are the performance and implementation results of the best model in classifying dry and wet coughs based on evaluation metrics including accuracy, precision, recall, F1-score, specificity, balanced accuracy, MCC, and the confusion matrix?

1.3 Research Objectives

Based on the problem statement, the objectives of this research are as follows:

1. To perform data selection, audio quality auditing, audio preprocessing, and Mel-Frequency Cepstral Coefficients (MFCC) feature extraction to produce consistent cough sound data ready for the classification process.
2. To design and implement an MFCC-based Capsule Network (CapsNet) model to classify dry and wet coughs on a balanced dataset.
3. To evaluate the performance of the best model in classifying dry and wet coughs using metrics such as accuracy, precision, recall, F1-score, specificity, balanced accuracy, MCC, and the confusion matrix, as well as to implement it into a web-based early screening system prototype.

1.4 Research Objectives

1. Academic Benefits
    a. To contribute to the advancement of science in Machine Learning and Deep Learning, particularly in the application of Mel-Frequency Cepstral Coefficients (MFCC) feature extraction methods and Capsule Network (CapsNet) architectures.

   b. To serve as a reference for future research in audio classification, particularly regarding healthcare applications or sound-based disease detection.
   c. To provide a basis for comparison with other classification methods (such as CNN or LSTM), thereby expanding academic studies regarding CapsNet performance in the audio domain.
2. Practical Benefits
   a. To develop a sound-based early screening system prototype that helps classify dry and wet coughs non-invasively, quickly, and practically.
   b. To help medical personnel or the general public obtain an initial indication of respiratory conditions, serving as a basis for seeking further medical evaluation.
   c. To serve as an initial prototype for developing artificial intelligence-based applications in healthcare, such as integration into mobile apps or health screening websites.
3. Social Benefits
   a. To help raise public awareness regarding the importance of early detection of respiratory diseases, particularly in areas with limited access to medical services.
   b. To encourage the use of digital technology and AI to support the healthcare sector, in line with HealthTech development in the digital era.
   c. In the long term, the findings of this research are expected to enhance public awareness regarding the importance of recognizing cough symptoms and seeking further medical evaluation when necessary.

1.5 Scope and Limitations

To ensure this research remains focused and well-directed, the following scope boundaries and limitations have been established:

1. The data used is limited to cough sound recordings sourced from a secondary data provider, specifically the public Kaggle dataset *"Cough Sound Dataset for Covid-19 Detection."* This study does not utilize primary data or direct cough recordings from volunteers.
2. The classified cough types are restricted to two main categories: dry cough (non-productive) and wet cough (productive). Other cough categories, such as chronic cough, coughing up blood (hemoptysis), or specific symptom variations, are not addressed in this research.
3. The final experimental dataset consists of a balanced, confident subset totaling 800 audio samples, comprising 400 dry cough samples and 400 wet cough samples.
4. The dataset is partitioned using the stratified split method with a ratio of 70:15:15, resulting in 560 training samples, 120 validation samples, and 120 testing samples. Each subset maintains a balanced distribution of dry and wet data.
5. Audio preprocessing steps include converting formats to WAV when necessary, converting stereo to mono, resampling to 16 kHz, trimming silence, adaptive energy-based cough event segmentation, and padding or cropping audio signals to a uniform duration of 3 seconds.
6. Audio feature extraction is performed exclusively using the Mel-Frequency Cepstral Coefficients (MFCC) method. Other methods, such as the Short-Time Fourier Transform (STFT), Linear Predictive Coding (LPC), or the Wavelet Transform, are not used.
7. The primary classification algorithm used is the Capsule Network (CapsNet). Other models, such as CNN, RNN, or SVM, are used only as baseline comparisons when necessary, with the main focus remaining on the implementation of CapsNet.
8. The scope of this research is limited to developing the cough type classification model and evaluating its performance. The classification results serve solely as an early screening indicator for respiratory conditions and are not intended to replace medical diagnoses made by professional healthcare providers.

9. Model evaluation is conducted using metrics including accuracy, precision, recall, F1-score, specificity, balanced accuracy, MCC, and the confusion matrix. Other evaluations, such as AUC-ROC analysis or clinical trials, are outside the scope of this study.