

学位論文要旨

(Summary of the Doctoral Dissertation)

学位論文題目 (Dissertation Title)	肺聴診音のコンピュータ支援診断に向けた深層学習モデルおよび異常検知手法の構築に関する研究
氏名(Name)	福永 亮佑

医療診断において、ある患者の状態を判断するための検査方法は多岐にわたり、検査結果を総合的に判断する必要がある。そこで、医師の診断の補助を目的とし、疾患に関連する症状や病変の解析を行うコンピュータ支援診断(Computer-Aided Diagnosis; CAD)技術が存在する。しかし、CADは画像診断に対しての試みが大部分であり、特に音声信号処理の分野においてはCAD技術の十分な検討がなされていない。そのため、人工知能(Artificial Intelligence; AI)を用いた音声信号処理技術の発展に期待が高まっている。CADを音声信号に対して適用することができれば、診断にかかる医師の負担を軽減できる。そこで、本論文では、肺疾患の診断方法の1つである聴診に対し、AIを用いた音声信号処理を応用することを考える。聴診音は、正常音や異常音について周波数帯域が重なることから統計的な識別が困難であり、さらに、診察環境の様々な要因によって複数種類のノイズが混入し得るため、古典的な音声信号処理では扱いが難しいデータである。近年の深層学習技術の進歩により、このような音声信号に対しても、AIモデルによる高精度な判別が可能になると期待できる。

ここで、肺聴診音のCAD技術を実現するためには、大きく3つの課題がある。1つ目の課題は、CAD技術としてどのようなタスク(分類タスク、異常検知タスクなど)を設定し、AIモデルに解かせるかである。AI技術を用いる場合、まず解くべきタスクの設定が重要であり、設定したタスクによって得られる精度や実用面の課題が異なる。2つ目の課題は、学習データの少なさである。聴診音の収集やラベリングは実務上の負担が大きいことに加え、異常音の発生は稀であるため、モデルの学習に十分なデータを用意することが難しい。そのため、少ない学習データに過剰適合しない、汎化性能の高いモデルが求められる。3つ目の課題は、識別に有効な肺聴診音の特徴量の選択である。学習データの少ない肺聴診音でCAD向けのAIモデルを構築する場合、学習効率の高さと識別精度の高さを両立したAIモデルが求められる。その実現のためには肺聴診音の前処理とAIアーキテクチャの適切な組み合わせによって、各タスクに重要な特徴が抽出できる方式が必要である。

本論文では、これらの課題に対する解決方法の提案及び検証を行った。第1章、第2章、第3章では、研究の背景や目的、本論文で用いる信号処理やAIモデル、使用するデータと評価指標についてそれぞれ説明する。

第4章では、学習データの少なさに対処することを目的として、メル周波数ケプストラム係数(Mel Frequency Cepstral Coefficient; MFCC)の拡張と事前学習を提案し、分類タスクの観点から2つ目と3つ目の課題の解決を図った。実験の結果より、2つ目の課題に対する事前学習の効果が確認された。また、ノイズ軽減の仕組みを持つアーキテクチャの性能が高い傾向があり、3つ目の課題について、周波数帯域やその時間変化の情報には識別上有効な特徴量が存在することが認められた。

第5章では、深層学習で、肺聴診音の識別において重要な特徴量をより効果的に処理することを目的として、転置MFCCと独自の分類モデルを提案し、3つ目の課題についてさらなる解決を図った。ここで、周波数帯域の情報が識別において特に有効であることや、その変化情報についても有用な特徴は含まれるなど、肺聴診音の特徴抽出において考慮すべき点を明らかにした。また、1つ目の課題である実用面について、分類モデルの予測結果の解釈性についても検討を行った。

第6章では、異常検知タスクに着目し、学習データの少なさに対処した、DAGMM(Deep Autoencoding Gaussian Mixture Model)とEfficient GAN(Efficient Generative Adversarial Network)の2種類の異常検知モデルを拡張した手法を提案し、2つ目の課題の解決を図りつつ、1つ目の課題である適切なタスク設定の検討を行った。実験では、DAGMMの拡張手法が高いAUC(Area Under the Curve)を達成し、異常検知モデルの性能が優れていることを示した。

様式7号（第12条，第31条関係）

（様式7号）（Format No.7）日本語版

第7章では，異常検知モデルの解釈性を高めることを目的として，位相的データ解析（Topological Data Analysis ; TDA）と Isolation Forest を組み合わせた異常検知手法について提案し，異常検知における1つ目の課題の解決を図った．提案した異常検知手法は，深層学習を用いた従来の異常検知モデルと同等の性能を持ち，高い解釈性を持つことが確認された．

第8章では，本論文全体で得られた結果を整理し，結論を示している．本論文の成果は，肺聴診音のCAD技術の実用化に向けた知見や，今後の精度向上の示唆を与えるものである．今後は，複数の聴診環境を跨いだ汎用的なモデルの構築や，診断支援の実用による医師の評価を得ることで，さらなる性能向上を図っていきたい

学 位 論 文 要 旨

(Summary of the Doctoral Dissertation)

学位論文題目 (Dissertation Title)	Study on the Construction of Deep Learning Models and Anomaly Detection Methods for Computer-Aided Diagnosis of Lung Sounds
氏 名(Name)	FUKUNAGA Ryosuke

In medical diagnosis, there are a variety of examination methods to assess a patient's condition, and the physician must make a comprehensive assessment based on the results. To assist physicians in making diagnoses, Computer-Aided Diagnosis (CAD) technologies have been developed to analyze disease symptoms and abnormalities. However, the majority of CAD applications have focused on imaging diagnostics, and CAD technology has not been sufficiently explored in the field of acoustic signal processing. In recent years, there is growing anticipation for advancements in acoustic signal processing technologies utilizing Artificial Intelligence (AI). If CAD technology for acoustic signals can be realized, it could potentially reduce the workload of physicians during diagnosis. Therefore, in this thesis, we explore the application of AI-based acoustic signal processing to auscultation, one of the diagnostic methods for lung diseases. Lung sounds is difficult to be classified by statistical analysis because normal and abnormal sounds overlap in frequency bands, and they are prone to various types of noise introduced by different diagnostic environments. It is difficult for traditional acoustic signal processing methods to handle such complex data. However, with recent developments in deep learning, it is anticipated that AI models will enable highly accurate classification of these acoustic signals.

In realizing CAD technology for lung sounds, there are three major challenges. The first challenge is defining which tasks, such as classification or anomaly detection, the AI model should solve. When applying AI, it is crucial to first determine the task to be addressed. As the task definition will influence both the accuracy and the problems to be considered for practical use. The second challenge is the small number of training data. In addition to the significant practical burden of physicians to collect and annotate lung sounds, the occurrence of abnormal sounds is rare, making it difficult to obtain enough data for model training. Therefore, there is a need for models with high generalization performance that can avoid overfitting to small datasets. The third challenge is the effective feature extraction from lung sounds for classification or detection. When constructing AI models for CAD with limited auscultation sound data, it is essential to balance high training efficiency with high classification accuracy. To achieve this, it is necessary to develop a method that can extract important features for each task by appropriately combining the preprocessing of lung sounds with AI architectures.

In this paper, we propose and validate solutions to these challenges. Chapters 1, 2, and 3 explain the background and objectives of the study, the signal processing techniques and AI models used in this paper, as well as the data and evaluation metrics employed.

In Chapter 4, we address the issue of limited training data by proposing the extension of Mel Frequency Cepstral Coefficients (MFCC) and pre-training techniques. These methods aim to solve the second and third challenges from the perspective of the classification task. The experimental results show the effectiveness of pre-training in addressing the second challenge. Additionally, architectures that incorporate noise reduction mechanisms show better performance, and with regard to the third challenge, it was found that features related to frequency bands and their temporal variations contain useful information for classification.

In Chapter 5, to more effectively generate the key features important for lung sound classification through deep learning, we propose the transposed MFCC and a custom classification model, further addressing the third challenge. This chapter clarifies that frequency band information is particularly effective for classification, and that temporal variations also contain valuable features that should be considered in feature extraction for lung sounds. Additionally, regarding the first challenge, which relates to practical application, we examine the interpretability of the predictions obtained by the classification model.

In Chapter 6, we focus on the anomaly detection task to address the issue of limited training data. We propose extended methods for two types of anomaly detection models: DAGMM (Deep Autoencoding Gaussian Mixture Model) and Efficient GAN (Efficient Generative Adversarial Network), aiming to solve the second challenge while also examining the appropriate task setting, which is the first challenge. In the experiments, the extended DAGMM method achieved a high AUC (Area Under the Curve), demonstrating the superior performance of the anomaly detection model.

In Chapter 7, to enhance the interpretability of anomaly detection models, we propose an anomaly detection method that combines Topological Data Analysis (TDA) with Isolation Forest, addressing the first challenge in anomaly detection. The proposed method demonstrated performance comparable to conventional deep learning-based anomaly detection models, while also offering high interpretability.

In Chapter 8, we summarize the results obtained throughout this study and present the conclusions. The findings of this research provide insights for the practical implementation of CAD technology for lung auscultation sounds and offer suggestions for future improvements in accuracy. Moving forward, we aim to further enhance performance by building a generalized model that can operate across multiple auscultation environments and by obtaining feedback from physicians through the practical use of the diagnostic support system.

学位論文審査の結果及び最終試験の結果報告書

山口大学大学院創成科学研究科

氏名	福永 亮佑
審査委員	主査：間普 真吾
	副査：中村 秀明
	副査：田村 慶信
	副査：藤田 悠介
	副査：佐村 俊和
論文題目	肺聴診音のコンピュータ支援診断に向けた深層学習モデルおよび異常検知手法の構築に関する研究 (Study on the Construction of Deep Learning Models and Anomaly Detection Methods for Computer-Aided Diagnosis of Lung Sounds)
【論文審査の結果及び最終試験の結果】 医師の診断補助を目的としたコンピュータ支援診断技術の研究が活発に行われているが、画像診断に対する試みがほとんどであり、特に音声信号処理においては十分な研究がされていない。そこで本論文では、医療診断方法の一つである聴診に着目し、人工知能 (AI) によって正常音と異常音を判別することを目的としている。しかし、医用音声信号は、データ収集の難しさ、周波数帯の重なりによる判別の難しさ、またノイズが含まれており、AI による高精度な判別のためには、これらの問題を解決することが必要であった。そこで、本論文では大きく 3 つの課題を設定し、その解決方法を提案している。①どのような AI タスク (分類, 異常検知) を設定することが判別に有用か, ②学習データの少なさをどうカバーするか, ③判別に有効な聴診音の特徴抽出方法は何か, である。 ①に対しては、分類タスクと異常検知タスクを設定し、AI 予測の解釈性も考慮した方式を提案, ②に対しては、分類モデルへの事前学習の導入, 少数データでも機能する異常検知モデルの応用と改良, ③に対しては、データの前処理の工夫, 聴診音解析に適した AI アーキテクチャの開発を行っている。 本論文の構成と内容は以下の通りである。 第 1 章では、研究の背景と目的, および論文の構成について述べている。 第 2 章では、関連研究について述べており、従来の音声信号処理法, 本論文で提案するモデルの構成要素となる、複数のニューラルネットワークアーキテクチャを説明したあと、これらを活用した分類モデルと異常検知モデルを主として説明している。 第 3 章では、使用した聴診音データセットの詳細な説明がなされ、分類および異常検知における評価指標に関して説明している。 第 4 章では、分類タスクを設定し、学習データの少なさに対応するために、複数のニューラルネットワークアーキテクチャに対して、独自の事前学習法を実装するとともに、データの前処理として実行されるメル周波数ケプストラム係数 (MFCC) の計算に対して、複数の	

抽出次元数の設定を行うこと、また MFCC の一次微分及び二次微分情報も活用する方式を提案している。上記を組合せた複数の提案方式を、219 症例からなる聴診音計測データに対して適用し、分類性能の向上を示したほか、提案方式間の性能比較も行っている。

第 5 章は、第 4 章と同様に分類タスクを設定しているが、特に時系列ニューラルネットワークの性能向上、および分類結果の解釈性に焦点を当てて研究を行っている。音声データからより多くの情報を抽出するため、転置 MFCC を導入し、時系列軸と周波数軸の両方でその変化を捉えられる仕組みとしている。また、Cross-encoding Transformer と呼ばれる、高い分類性能と解釈性を有するモデルを提案している。4 章と同様の実験により、提案方式による分類性能の向上を示した。しかし、分類において重視されたデータの時刻を可視化したものの、根拠の説明として分かりにくい箇所があり、改善が必要である結果となった。

第 6 章では、異常検知タスクを設定し、2 種類の異常検知モデル (Deep Autoencoding Gaussian Mixture Model (DAGMM) 及び Efficient GAN の改良) を提案している。改良のポイントは、第 4 章で良い性能を示したニューラルネットワークアーキテクチャを導入すること、及び Efficient GAN に混合ガウスモデルを統合することで、正常音と異常音の特徴の境界を明瞭にする学習を可能にすることである。実験の結果、DAGMM、Efficient GAN ともに性能向上が見られ、とくに DAGMM の改良が最も良い性能を示した。

第 7 章でも、異常検知タスクを設定しているが、第 6 章までと異なり、深層学習を用いない Isolation Forest と呼ばれる手法をベースに、解釈性を重視した異常検知モデルを提案している。具体的には、解釈性を持たせ、さらに深層学習と同等の検知性能を保持するため、トポロジー特徴量、相関係数、音声 IDF を導入した Isolation Forest を構築している。実験の結果、深層学習ベースの手法と同等の検知性能を示しつつ、判定の根拠となったデータの時刻をより明確に示すことが可能になった。

第 8 章では、本研究の成果をまとめ、結論を述べている。

公聴会には 33 名の参加があり、活発な議論が行われた。主な質疑内容として、

- 本研究における AI の出力の解釈性の限界について
 - 音声 IDF を提案しているが、音声信号の時間変化を捉えた予測のための役割は何か
 - AI による診断のリアルタイム性
 - 実環境では様々なノイズが入るが診断性能への影響はどの程度あるか
- 等があり、いずれの質問に対しても申請者からの確かな回答がなされた。

以上より、本研究は独創性、信頼性、有効性、実用性とも優れ、博士 (工学) の論文に十分値するものと判断した。

論文内容及び審査会、公聴会での質問に対する応答などから、最終試験は合格とした。

なお、主要な関連論文の発表状況は以下の通りである。(関連論文計 5 件)

1. 若本亮佑, 間普真吾, 木戸尚治, 呉本堯, 事前学習つき深層ニューラルネットワークを用いた肺聴診音識別—CNN, LSTM, 畳み込み LSTM の性能比較—, 電気学会論文誌 C, 140 巻 12 号, pp. 1402-1409, 2020 年 12 月発行
2. 若本亮佑, 間普真吾, 木戸尚治, 呉本堯, 深層ニューラルネットワークを用いた肺聴診音の異常検知— DAGMM, Efficient GAN の性能比較と改良—, 電気学会論文誌 C, 142 巻 12 号, pp. 1328-1335, 2022 年 12 月発行
3. Ryosuke Wakamoto and Shingo Mabuchi, Effective Feature Extraction from Lung Sounds Using Time-Series Neural Networks, Proceedings of SICE Festival with Annual Conference 2024, pp. 846-851, 2024 年 8 月発行