JAMES COOK UNIVERSITY

DOCTORAL THESIS

Health Social Network of Things: Towards Integrated Autonomous Online Community Healthcare

---

Author: Yi Huang

A thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy (Information Technology)

in the

James Cook University Singapore

Dec 31, 2023

1 Declaration Of Authorship

I, Yi Huang, declare that this thesis titled *Health Social Network of Things: Towards Integrated Autonomous Online Community Healthcare* and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree in this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. Except for such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: 31th Dec 2023 _____

2 Acknowledgments

I would like to express my deepest gratitude and appreciation to my esteemed advisors, Dr. Insu Song and Prof. Ickjai Lee, for their invaluable guidance, unwavering support, and profound knowledge throughout my doctoral journey. Their expertise, dedication, and mentorship have been instrumental in shaping the outcome of my research, and I am truly honoured to have had the opportunity to work under their supervision.

I would like to extend my heartfelt thanks to Dr. George Jacobs for his invaluable coaching and support during the drafting of my publications. His meticulous attention to detail, linguistic expertise, and constructive feedback have significantly enhanced the clarity and effectiveness of my research communications. I am truly grateful for his generosity in sharing his time and expertise.

I am also indebted to Dr. Denise and the entire university staff for their constant support, administrative guidance, and valuable advice. Their professionalism, responsiveness, and willingness to assist have been instrumental in ensuring a smooth and productive academic journey. The dedicated team at JCU Singapore has consistently demonstrated professionalism and a genuine commitment to supporting students, which has greatly contributed to my overall experience.

Lastly, I would like to express my deepest appreciation to my family for their unwavering love and encouragement throughout this challenging endeavour. Their constant support, understanding, and belief in my abilities have been a constant source of motivation, strength, and inspiration. I am truly blessed to have such a loving and supportive family, and I dedicate this achievement to them.

To all those mentioned above, and to countless others who have contributed to my growth and success, I offer my sincerest appreciation and gratitude. Your guidance, encouragement, and support have been invaluable, and I am forever grateful for your contributions to my academic and personal development.

3 Statement of the Contribution of Others

Nature of Assistance	Contribution	Name
Supervision support	Primary, secondary and mentor supervision respectively	Dr Insu Song, and Prof Ickjai Lee
Research support	Drafting, paper/thesis design, review and correctness check	Dr Insu Song
Proofreading support	Proofreading for english writing of this thesis	Dr. John Cokley

4 Publications Associated with Chapters

Chapter 3.1

Huang, Y., & Song, I. (2018). *A Better Online Method of Heart Diagnosis*. Paper presented at the Proceedings of the 3rd International Conference on Biomedical Signal and Image Processing, Seoul, Republic of Korea.

Chapter 3.2

Huang, Y., & Song, I. (2019). *Indexing Biosignal for Integrated Health Social Networks.* Paper presented at the ICBBE 2019, China, Shanghai.

Huang, Y., & Song, I. (2024). Indexing ECG for Integrated Health Social Networks. *SN Computer Science, 2024.*

Chapter 3.3

Huang, Y., & Song, I. (2022, 24-26 June 2022). *Cardio Vec: Searching Heart Health Information Using ECG Signals.* Paper presented at the 2022 7th International Conference on Computational Intelligence and Applications (ICCIA).

Chapter 4.2

Huang, Y., & Song, I. (2022, 8-10 July 2022). *PhysioVec: A Multi-stage Deep-Learning Framework for Searching Online Health Information with Breath Sound.* Paper presented at the 2022 IEEE 5th International Conference on Big Data and Artificial Intelligence (BDAI).

Chapter 5.2

Song, I., Huang, Y., Koh, T. H. H. G., & Rajadurai, V. S. (2021, 2021//). *Pervasive Monitoring of Gastrointestinal Health of Newborn Babies.* Paper presented at the PRICAI 2021: Trends in Artificial Intelligence, Cham.

Chapter 5.3

Huang, Y., & Song, I. (2022, 24-26 June 2022). *PhysioVec: IoT Biosignal Based Search Engine for Gastrointestinal Health.* Paper presented at the 2022 7th International Conference on Computational Intelligence and Applications (ICCIA).

5 Abstract

This thesis explores the integration of Health Social Networks (HSN) with the Internet of Things (IoT) to enhance the accessibility and accuracy of health information for patients and caregivers. HSNs refers to the collection of patient-driven communities that provide rich information to patients and caregivers. On the other hand, Health IoT enables remote monitoring of patients to provide non-invasive, objective, and affordable diagnostic capabilities. By combining HSNs and Health IoT, this research aims to facilitate quick and easy access to relevant health information while enabling pervasive remote monitoring for improved healthcare outcomes.

In this thesis, the challenges associated with integrating HSNs, and Health IoT are identified. The integrated HSN focuses on the application areas of heart disease, lung disease and bowel disease.

For heart diseases, the first challenge is finding similar patient groups in HSNs using biosignals collected through Health IoT as users often struggle to locate relevant communities for specific diseases in HSNs. Users often lack expertise in effectively describing their condition, making it challenging to find appropriate communities within HSNs. The second challenge is converting ECG (electrocardiogram) results into keywords that allow direct comparison with textual data for HSN search. Conventional ECG classification methods can only predict predefined diseases and abnormalities from their training dataset, which restricts their ability to cover the wide range of possibilities that have not been explicitly trained for. The third challenge is enabling semantic search of text-based contents from HSNs using ECG from Health IoT. This challenge involves developing methods to convert biosignal data into clinically interpretable descriptions as keywords to search relevant health information from HSNs.

For respiratory diseases, the first challenge is developing an efficient deep learning model that effectively combines the strengths of Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers for the analysis of breath sound. Due to gradient exploding or vanishing, RNNs have limitations to model diverse spatial relations and long sequences. CNNs have advantages to compose lower-level elements into higher-level entities, making them efficient for signal processing tasks. On the other hand, CNNs face constraints in modelling features with varying distances due to their reliance on fixed aggregation weights in its convolutional kernel for spatially neighbouring input features. The second challenge is enabling semantic search of text-based contents from HSNs using breath sound from Health IoT. This challenge involves developing methods to convert biosignal data into clinically interpretable descriptions as keywords to search relevant health information from HSNs.

For bowel diseases, the first challenge is collecting bowel sounds for detecting bowel diseases in newborns within the Neonatal Intensive Care Unit (NICU). Specifically, affordable methods for collecting bowel sounds are lacking. The second challenge is using deep learning models to accurately segment bowel sounds. The irregular pattern of bowel sounds poses difficulties in locating

and detecting bowel activities. As a result, conventional deep learning models often face challenges in precisely segmenting bowel sounds, thus limiting their effectiveness in diagnosing gastrointestinal problems. The third challenge is enabling semantic search of text-based contents from HSNs using bowel sounds from Health IoT. This challenge involves developing methods to convert biosignal data into clinically interpretable descriptions as keywords to search relevant health information from HSNs.

The thesis aims to develop an integrated HSN framework that encompasses heart disease, lung disease and bowel disease, providing comprehensive and efficient healthcare solutions.

For cardiovascular disease, the first aim is to enable direct biosignal comparison through Health IoT, facilitating the identification of similar patient groups in HSNs and overcoming challenges in finding relevant communities for specific diseases without requiring users to search on their own. The second aim is to enable effective semantic search of cardiovascular disease-related health information from HSNs, expanding the coverage of healthcare data possibilities beyond predefined diseases and abnormalities. The third aim is to convert the ECG into interpretable clinical measurements that describe physical measurement abnormalities, to provide keywords for semantic search within HSNs.

For respiratory disease, the first aim is to develop an efficient deep learning model that combines CNNs, RNNs and Transformers to analyse breath sounds, addressing limitations in modelling diverse spatial relations and handling long sequences. The second aim is to convert the breath sounds into interpretable clinical measurements that describe physical measurement abnormalities, to provide keywords for semantic search within HSNs.

For gastrointestinal disease, the first aim of this research is to develop cost-effective methods for collecting bowel sounds and extracting important information about bowel health conditions in the NICU for the purpose of detecting bowel diseases in newborns. The second aim is to develop deep learning models that accurately segment and detect bowel sounds to improve the diagnosis of gastrointestinal problems, despite the challenges posed by the irregular pattern of bowel sounds. The third aim is to convert the bowel sounds into interpretable clinical measurements that describe physical measurement abnormalities, to provide keywords for semantic search within HSNs.

To achieve the aims, the thesis begins with a comprehensive survey of the HSN and Health IoT. This survey identifies gaps and opportunities in the field. Subsequently, the following approaches are developed to address the challenges and aims identified in the survey.

For cardiovascular diseases, the first approach diagnoses heart diseases by utilising the Gaussian Hamming Distance, a similarity comparison method. By comparing the similarity between heart sounds of individuals with other patients from online support groups, this approach enables accurate searching of online support groups for certain diseases. The second approach develops a Keyword-based Integrated HSN of Things (KIHoT), which predicts keywords embedding from ECG signals.

These keywords are utilised to search HSNs for content with similar keywords, facilitating the prediction and identification of heart conditions. The third approach develops CardioVec, a system that utilises statistical analysis of specific physiological measurements extracted from the ECG. These measurements are then processed to generate human-readable keywords. These keywords can be used to search for relevant information or connect with other users on HSN.

For respiratory diseases, the first approach proposes a deep learning model called Recurrent Local Relation Transformer Encoder Classifier (ReLATEC) that utilises local attention and recurrent structure to diagnose respiratory diseases from breath sounds collected with mobile devices, that combines CNNs, RNNs and Transformers to analyse breath sounds. The second approach develops PhysioVec to enable semantic searching of breath sounds. PhysioVec first utilises semantic segmentation techniques to extract breath sound patterns. PhysioVec then processes these biosignals to extract specific physiological measurements. Through statistical analysis of these measurements, PhysioVec generates human-readable keywords that are used for semantic searching within HSN, enabling the retrieval of relevant information. The third approach involves using a miniature Bowel Sounds Sensor (BoSS) to collect bowel sounds in the NICU. These bowel sounds are then processed to extract physiological measurements. The subsequent analysis of these measurements offers valuable information for the diagnosis and monitoring of bowel conditions in newborns.

For gastrointestinal disease, the first approach collects bowel sounds of newborn infants using BoSS. The second approach trains ReLATEC with novel weak annotations to accurately segment and detect bowel sounds, improving the diagnosis of gastrointestinal problems in newborns. The third approach develops PhysioVec to enable semantic searching of bowel sounds. PhysioVec first utilises semantic segmentation techniques to bowel breath sound patterns. PhysioVec then processes these biosignals to extract specific physiological measurements. Through statistical analysis of these measurements, PhysioVec generates human-readable keywords that are used for semantic searching within HSN, enabling the retrieval of relevant information.

The contributions of the thesis are listed as follows:

For cardiovascular disease, the first contribution is the introduction of Gaussian Hamming Distance to identify similar heart condition groups within HSNs. Through the comparison of heart sounds among users in the HSNs, this approach enables the identification of individuals with similar heart sounds. The second contribution is the proposal of KIHoT to enable direct comparison of ECG signals with textual data for comprehensive access to health information. The third contribution is the development of CardioVec, which enables semantic search in HSN. CardioVec utilises clinically interpretable descriptions derived from ECG to improve search efficiency and enhance the trustworthiness of the reasoning process.

For respiratory disease, the first contribution involves the design of the ReLATEC for precise diagnosis of respiratory diseases using breath sounds captured from mobile devices. This

contribution enhances the accuracy of respiratory disease diagnosis by combining the strengths and weaknesses of CNN, RNN and Transformer models. The second contribution is the development of PhysioVec to enable semantic search of text-based content in HSNs using clinically interpretable descriptions derived from breath sounds, improving search efficiency.

For gastrointestinal disease, the first contribution is the development of a miniature BoSS for detecting bowel diseases in newborns, improving early detection in NICUs. Physiological measurements from bowel sounds are extracted and identified for detecting gastrointestinal issues in newborns. This study is the first to collect and analyse bowel sounds of term newborns, addressing the lack of continuous monitoring methods. The second contribution is enhancing the ReLATEC model by training it with bowel-sound data to improve the diagnosis of gastrointestinal issues in newborns. By providing interpretable clinical formats, this contribution enhances the understanding and extraction of meaningful information from biosignals, improving healthcare decision-making processes. The third contribution is the development of PhysioVec to enable semantic search of text-based content in HSNs using clinically interpretable descriptions derived from bowel sounds, improving search efficiency.

Keywords: Biomedical engineering, bio informatics, data mining, noise reduction, diagnosis using mobile devices.

6 Table Of Contents

7 List of Figures

8 List of Tables

9 List of Abbreviations

Concepts:

Health Social Network (HSN)

Internet of Things (IoT)

Artificial Intelligence (AI)

Neural Networks:

Multiple Layer Perceptron (MLP)

Convolutional Neural Network (CNN)

Recurrent Neural Network (RNN)

Convolutional Recurrent Neural Network (CRNN)

Long-Short-Term Memory (LSTM)

Bidirectional LSTM (BiLSTM)

Recurrent Local Attention Transformer Encoder Classifier (ReLATEC)

Local Recurrent Transformer (LRT)

Conditional Random Fields (CRF)

Gated Recurrent Unit (GRU)

Universal Transformer (UT)

Local Relation Network (LR-Net)

Gaussian Error Linear Units (GeLU)

Bidirectional Encoder Representations from Transformers (BERT)

Sentence-BERT (SBERT)

ALBERT (A Lite BERT)

Large Language Models (LLMs)

Retrieval-Augmented Generation (RAG)

Stochastic Gradient Descent (SGD)

Adam with Weight Decay (AdamW)

Features:

Short-Time Fourier Transform (STFT)

Doubly Labelled Water (DLW)

Fast Fourier Transform (FFT)

Discrete Wavelet Transform (DWT)

Continuous Wavelet Transform (CWT)

Wavelet Packets (WP)

Mel-Frequency Cepstral Coefficients (MFCC)

Power Spectral Density (PSD)
Hilbert Huang Transform (HHT)
Path Long Entropy (PLE)
S-transform (ST)
Wigner-Ville distribution (WVD)
Choi-Williams distribution (CWD)

Machine Learning:

Gaussian Hamming Distance (GHD)
K-Nearest Neighbour (KNN)
Support Vector Machine (SVM)
Univariate Autoregressive (UAR)
Multivariate Autoregressive Models (MAR)
Bag of Words (BOW)
Regions of Interests (ROIs)
Natural Language Processing (NLP)
Computer Vision (CV)
Area under Curve (AUC)
Receiver Operator Characteristic (ROC)
Gaussian Raw Data Time Series (GTS)
Gaussian Time Series with Absolute Value (GATS)
Gaussian Time Series with Absolute Value and Pooling (GPATS)
Gaussian MFCC Time Series (GMTS)
Leave-One-Out-Validation (LOOV)
Keyword based Integrated HSN of Things (KIHoT)
Sums of The Word Embedding Vectors (SOWE)
Continuous BOW (CBOW)
Semantic Text Similarity (STS)
Intersection over Union (IoU)
Multivariate Radial-Basis Logistic Interpreter (MLI)
Mean Squared Error (MSE)
Root Mean Square Error (RMSE)
Linear Predictive Coding (LPC)
Retrieval-Augmented Generation (RAG)

Medical:

Myocardial Infarction (MI)
Intrauterine Growth Restriction (IUGR)
Large for Gestational Age (LGA)
Electrocardiogram (ECG)
Adverse Drug Reactions (ADRs)
PCG (Phonocardiogram, contain both ECG and heart sound)
Online health community (OHC)
Post Traumatic Stress Disorder (PTSD)
Attention Deficit Hyperactivity Disorder (ADHD)
Necrotising Enterocolitis (NEC)
World Health Organization (WHO)
Neonatal Intensive Care Unit (NICU)
Bowel Sounds Sensor (BoSS)
Reverse Transcription Polymerase Chain Reaction (RT-PCR)
Automatic Computer Diagnosis (CAD)
Seasonal Affective Disorder (SAD)
Electronic Health Records (EHR)
Coronavirus Disease 2019 (COVID-19)
Antepartum haemorrhage (APH)
Twin Transfusion Syndrome (TTS)
American Association for the Study of Liver Diseases (AASLD)

Datasets:

PTBDB (A physio dataset for ECG)
CADEC (A dataset called CSIRO Adverse Drug Event Corpus)
KK hospital
NS dataset
DK dataset

1 Introduction

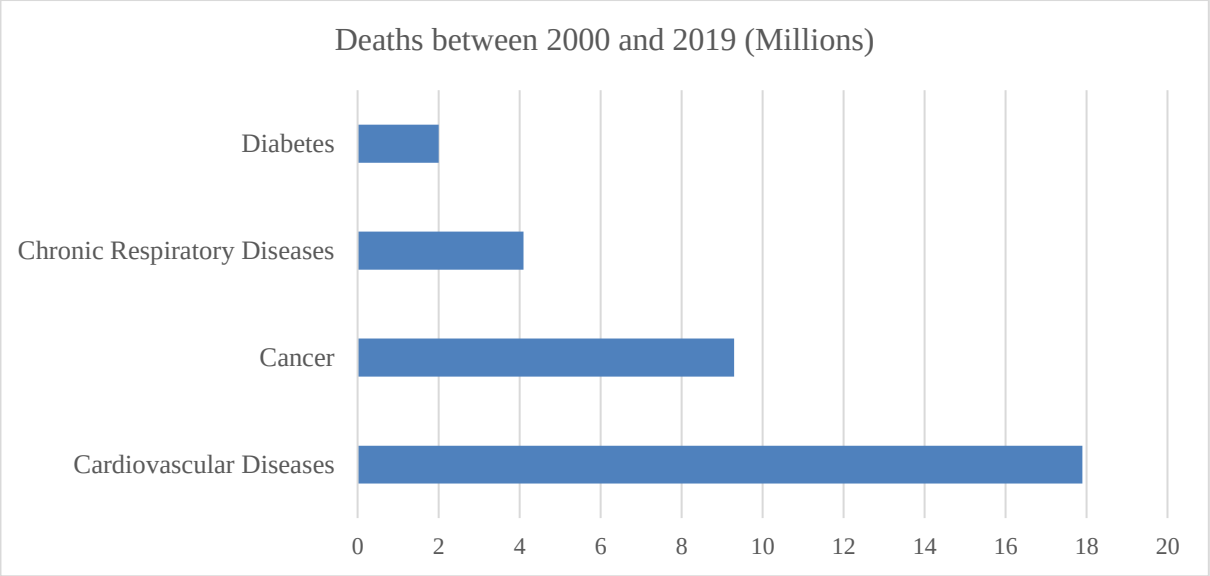

Figure 1.1 ***The four main types of NCDs worldwide in 2023.***

Noncommunicable diseases (NCDs) have become a leading cause of death and disability worldwide, resulting in a significant number of deaths and disability-adjusted life years. Between 2000 and 2019, the four major NCDs, namely cardiovascular disease, cancer, chronic respiratory disease, and diabetes, accounted for a significant number of deaths. Cardiovascular disease alone caused 17.9 million deaths, followed by cancer with 9.3 million deaths, chronic respiratory disease with 4.1 million deaths, and diabetes with 2.0 million deaths as shown in Figure 1.1.

Progress has been made in all regions since 2000, with a global improvement of 22% in the likelihood of dying from NCDs before the age of 70 (World Health Organization, 2023). Although progress has been made, these diseases remained leading causes of death in the United States in 2021, with heart disease, cancer, Coronavirus Disease 2019 (COVID-19), and other factors contributing significantly. In the United States, the leading causes of death in 2023 were heart disease (695,547 deaths), cancer (605,213 deaths), COVID-19 (416,893 deaths), unintentional injuries (224,935 deaths), stroke (162,890 deaths), chronic lower respiratory diseases (142,342 deaths), Alzheimer's disease (119,399 deaths), diabetes (103,294 deaths), chronic liver disease and cirrhosis (56,585 deaths), and nephritis, nephrotic syndrome, and nephrosis (54,358 deaths) (Xu, Murphy, Kochanek, & Arias, 2022). Similarly, in Australia in 2023 respiratory diseases accounted for 9.0% of all doctor-certified deaths, with a total of 14,289 deaths, while heart-related conditions contributed to 14.7% of the overall mortality rate, with a total of 23,380 deaths (Australian Bureau of Statistics, 2024).

Continuous care and monitoring are necessary to prevent deaths caused by common chronic diseases such as heart disease. However, the rising cost of healthcare in an ageing society poses a significant challenge (Song & Marsh, 2012). For instance, in the US, the training cost for general participants exceeds US$300,000 (Vong & Song, 2015). Health Social Networks (HSN) and Internet

of Things (IoT) provide solutions for remote diagnosis to enable accessible remote diagnosis. HSNs allow individuals to connect, share health-related experiences and provide peer support, while IoT enables the passive and continuous collection of physiological data using networked medical sensors.

HSNs are patient-driven communities that provide rich information to patients and caregivers (Song & Marsh, 2012). While medical professionals and offline resources are the foundation of healthcare, people still need to share knowledge on social networks. HSNs have become an important source of information for patients and physicians to make informed decisions. In recent years, many automatic recommendation systems for social networks have been developed (Song & Marsh, 2012). For example, 20% of adults have found similar health problems from others online, and most of them have benefited from HSNs (Fox, 2011). On the other hand, the IoT provides a potential solution for objectively describing health conditions with clinical measurements. IoT sensors could collect bowel sounds to provide pervasive, low-cost and autonomous remote gastrointestinal health monitoring.

Although HSNs and Health IoT aim to improve accessibility and reduce healthcare costs, they have distinct advantages and limitations. HSNs face usability challenges as users encounter difficulties in finding relevant information. Patients often lack expertise in selecting appropriate keywords, leading to inconsistencies and misinformation in search results (Elhadad, Zhang, Driscoll, & Brody, 2014).

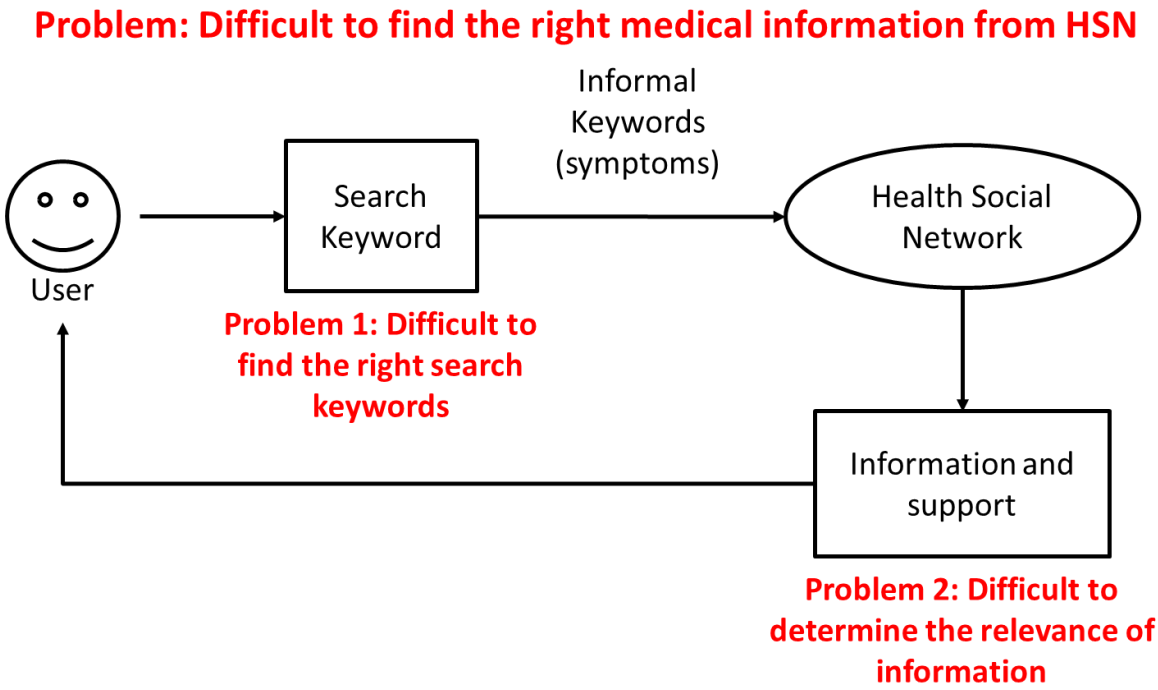

Figure 1.2 ***Common usage of HSNs.***

Figure 1.2 illustrates the common approach of HSN-based health monitoring, which involves users searching for relevant information amid many unrelated posts and articles. The substantial volume of data available in HSNs poses the most significant barrier for users seeking relevant information, as they must navigate a vast amount of data to find the desired content. HSNs rely on user-entered keywords to retrieve the desired information, but users' lack of expertise in selecting appropriate keywords can impact the accuracy and relevance of the search results (Alimova & Tutubalina, 2017; Miftahutdinov, Tropsha, & Tutubalina, 2017).

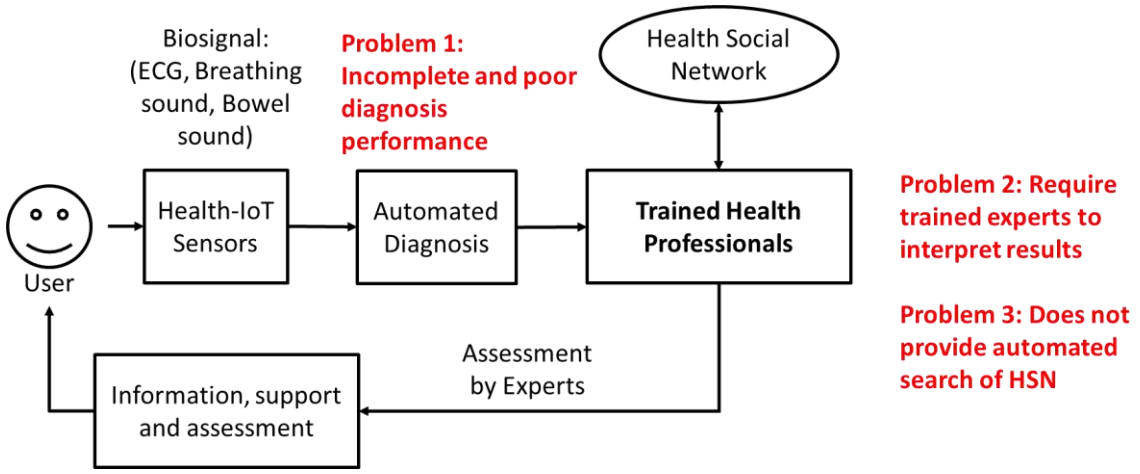

Figure 1.3 ***Common usage of IoT.***

The current Health IoT faces challenges in providing automated search capabilities for relevant health information due to the absence of interpretable biosignal pattern analysis. This poses a challenge to effective searches for health-related information within HSNs, necessitating manual interpretation and monitoring by trained medical professionals (Manogaran et al., 2018). Figure 1.3 illustrates that the absence of interpretable biosignal pattern analysis in existing IoT methods still requires experts to manually interpret and monitor biosignals, leading to costs and time consumption.

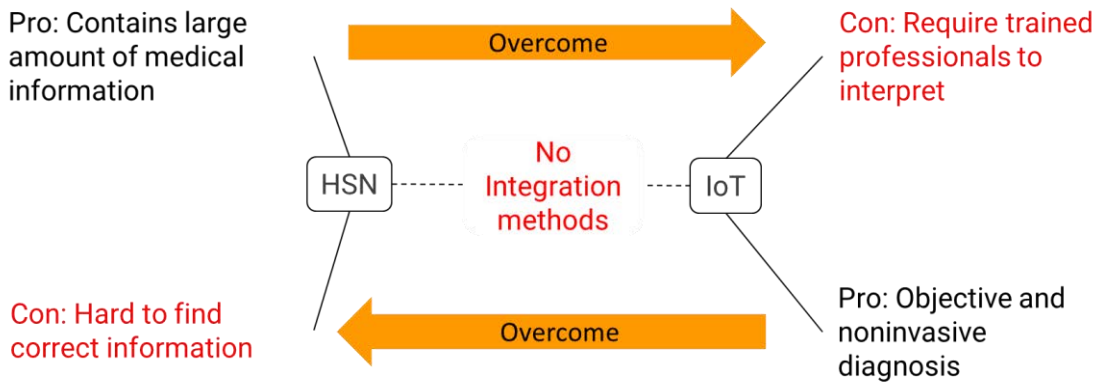

Figure 1.4 ***Research Problem: Integration of HSNs and IoT.***

Integrating HSNs with Health IoT technologies shows potential for advancing healthcare by facilitating low-cost, autonomous, and effective monitoring and management. However, despite our best efforts, we could not find an existing approach that integrates HSNs and Health IoT for this problem (Figure 1.4).

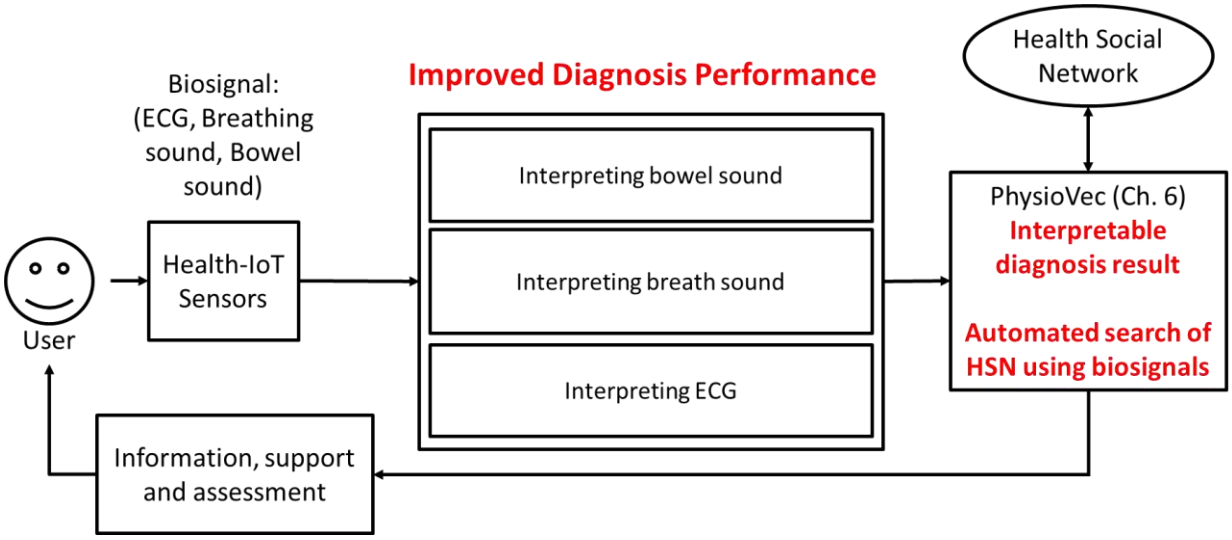

Figure 1.5 ***Common usage of our proposed integrated HSNs.***

The proposed integrated HSN aims to enable patients to access medical information and care without requiring expert medical knowledge, while also reducing unnecessary human intervention. This approach significantly reduces costs and improves access to medical care. Figure 1.5 illustrates the overall process of the proposed integrated HSNs. First, the physiological measurements of biosignals are extracted from the patient's body. After that, human-readable clinical descriptions are generated from physiological measurements. Finally, the clinical descriptions are used as search terms to retrieve relevant information from the internet.

1.1 Research Problems

The thesis aims to address the challenges of integrating HSN and Health IoT devices by focusing on three main application areas: heart disease, respiratory disease, and bowel disease. The challenges identified in each area are approached cohesively to solve specific problems related to each health condition.

For heart disease, the first challenge is utilising biosignal analysis to help individuals find online support groups that cater for their specific conditions. HSNs face challenges in accurately matching patients with similar conditions due to the informal nature of medical descriptions and the presence of spelling errors (Elhadad et al., 2014). Many users encounter difficulties in independently finding suitable online support groups due to their limited expertise in describing and locating relevant communities. The second challenge involves the effective utilisation of electrocardiogram (ECG) signals to improve the performance of disease prediction from ECG. Conventional classification

approaches currently in use have limitations. These approaches can only predict a predefined set of diseases and abnormalities, thereby failing to encompass the wide range of possibilities. As a result, this limitation hampers the comprehensive search for information within HSNs. The third challenge is providing automated search capabilities for relevant health information due to the absence of interpretable ECG pattern analysis. This absence hinders the effectiveness of searches for health-related information for cardiovascular disease within HSNs. Existing ECG diagnostic methods still require experts to interpret and monitor ECG data at the end, which is both costly and time-consuming (Manogaran et al., 2018).

In the case of respiratory disease, the first challenge pertains to the absence of an efficient deep learning model that effectively combines the strengths of Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNNs), and Transformers for the analysis of breath sounds. Conventional deep learning methods, such as CNN and RNN, have limitations in effectively modelling varying spatial relations and long sequences that involve long or varying distances (Vaswani et al., 2017). RNN possesses certain limitations when it comes to effectively modelling diverse spatial relations and handling long sequences, primarily due to issues such as gradient exploding or vanishing. Similarly, CNNs encounter limitations when dealing with features that have varying distances, as they rely on fixed aggregation weights for spatially neighbouring input features, thereby posing challenges in such scenarios (Hu, Zhang, Xie, & Lin, 2019). The second challenge is that current approaches lack interpretable breath pattern abnormality analysis for further understanding of respiratory conditions. This lack hinders the effectiveness of searches for health-related information within HSNs, thus necessitating manual interpretation and monitoring by trained medical professionals (Manogaran et al., 2018).

In the context of bowel disease, the first challenge is the lack of affordable methods for collecting biosignals, which poses challenges in the detection of bowel diseases in newborns within the Neonatal Intensive Care Unit (NICU). Currently, there is no known method for continuous monitoring of gastrointestinal health in term newborn infants at the NICU. In addition, locating and detecting intestinal sounds from digital recordings remains a challenge due to the irregular patterns and occurrence of these sounds (Ranta, Louis-Dorr, Heinrich, Wolf, & Guillemin, 2004; Sazonov et al., 2010). The nonstationary characteristics of bowel sounds, characterised by duration and amplitude variations, make it difficult to recognise and identify bowel sounds, particularly for physicians (Hadjileontiadis & Rekanos, 2003). The second challenge is the absence of known approaches that can provide real-time, visual monitoring of bowel functions in newborn infants. Currently, doctors diagnose bowel diseases in infants manually using a stethoscope, which can be uncomfortable and invasive for young children (Y. Huang & Song, 2019; Y. Huang, Song, Rana, & Koh, 2017; I. Song, 2015). This manual process of annotating and analysing bowel sounds is time-consuming, particularly in the presence of noise. Even experienced doctors struggle to identify only a few sporadic bowel

events when manually analysing bowel sounds (Ranta et al., 2004). Therefore, there is a need for automated bowel sound detection. The third challenge is that current automated bowel sound analysis approaches lack interpretable bowel sound pattern analysis for a deeper understanding of bowel conditions. This limitation hinders effective searches for health-related information within HSNs, thereby necessitating manual interpretation and monitoring by trained medical professionals (Manogaran et al., 2018). As a result, the current approach still relies on the manual interpretation of bowel sounds and the manual search for information on gastrointestinal health within HSNs.

1.2 Research Questions

The following are the research question statements of the problems discussed.

$RQ_{1.1}$: How can biosignal analysis be utilised to accurately match individuals with specific health conditions to suitable online support groups within HSNs by improving the performance for similarity measures of heart sounds? This research question is solved in Chapter 3 Section 3.1.

$RQ_{1.2}$: What approaches can be developed to improve the performance of disease prediction from ECG? This research question is solved in Chapter 3 Section 3.2.

$RQ_{1.3}$: How can interpretable ECG pattern analysis be integrated into direct search capabilities for cardiovascular disease-related health information within HSNs? This research question is solved in Chapter 3 Section3.3.

$RQ_{2.1}$: How can a deep learning model be developed that effectively combines CNN, RNN and Transformers to analyse breath sounds, overcoming the limitations of conventional methods? This research question is solved in Chapter 4 Section 4.1.

$RQ_{2.2}$: What approaches can be employed to enable interpretable breath pattern abnormality analysis for a better understanding of respiratory conditions, thereby enhancing the effectiveness of searches for health-related information within HSN? This research question is solved in Chapter 4 Section 4.2.

$RQ_{3.1}$: What affordable methods can be developed for the continuous monitoring of gastrointestinal health in newborns within the NICU, specifically focusing on the detection of bowel diseases? This research question is solved in Chapter 4 Section 4.1.

$RQ_{3.2}$: How can real-time, non-invasive visual monitoring of bowel functions be achieved in newborn infants, eliminating the need for manual diagnosis using a stethoscope? This research question is solved in Chapter 4 Section 4.2.

$RQ_{3.3}$: What approaches can be implemented to enable interpretable bowel sound pattern analysis for a deeper understanding of bowel conditions, enhancing the effectiveness of searches for health-related information within HSNs and reducing the reliance on manual interpretation by medical professionals? This research question is solved in Chapter 4 Section 4.3.

1.3 Hypotheses

The following are the hypotheses of research question statements.

- $H_{1.1}$: Using Gaussian Hamming distance over Euclidian distance improves performance for similarity measure of heart sounds.
    - Independent variables:
        - Similarity measure method: Gaussian hamming distance vs Euclidian distance.
    - Dependent variable:
        - performance for similarity measure of heart sounds.
- $H_{1.2}$: Training deep learning models using keyword embedding of disease description and ECG improves the performance of disease prediction from ECG.
    - Independent variables:
        - Training deep learning models using keyword embedding of disease description and ECG.
    - Dependent variable:
        - Performance of disease prediction from ECG.
- $H_{1.3}$: Interpretable ECG pattern analysis enables effective semantical search of cardiovascular textual disease-related health information from HSNs.
    - Independent variables:
        - Analysis ECG: Interpretable ECG pattern analysis.
    - Dependent variable:
        - The effective semantical search of textual cardiovascular disease-related health information from HSNs.
- $H_{2.1}$: Using a Transformer instead of a Convolutional Recurrent Neural Network (CRNN), local attention over global attention in Transformer, and recurrent Transformer block instead of independent Transformer blocks improves the performance of breath sound diagnosis compared with conventional deep learning approaches.
    - Independent variables:
        - Using a Transformer instead of a CRNN
        - Using local attention over global attention in Transformer
        - Using a recurrent Transformer block instead of independent Transformer blocks
    - Dependent variable:
        - Performance of breath sounds diagnosis.
- $H_{2.2}$: Interpretable breath pattern abnormality analysis enables effective semantical search capabilities for respiratory disease-related health information within HSNs.
    - Independent variables:
        - Analysis of breath sounds: Interpretable breath sound pattern analysis.
    - Dependent variable:
        - The effective semantical search of textual respiratory disease-related health information from HSNs.
- $H_{3.1}$: Bowel sounds collected from a compact bowel sound sensor provides significant information about bowel health conditions.
    - Independent variables:
        - Collect bowel sounds from a compact bowel sound sensor.
    - Dependent variable:
        - Significant differences in bowel sounds were collected from compact sensors of sick and healthy infants.

- $H_{3.2}$: Using the Transformer with localised attention instead of global attention and recurrent Transformer block instead of independent Transformer blocks improves performance for bowel sound location prediction.
    - Independent variables:
        - using a Transformer instead of a CRNN
        - attention method: Local vs global
        - parameter sharing: Recurrent Transformer block vs independent Transformer blocks
    - Dependent variable:
        - performance for bowel sound location prediction.
- $H_{3.2}$: Interpretable bowel pattern abnormality analysis enables effective semantical search capabilities for bowel disease-related health information within HSNs.
    - Independent variables:
        - Analysis of bowel sounds: Interpretable bowel sound pattern analysis.
    - Dependent variable:
        - The effective semantical search of textual bowel disease-related health information from HSNs.

1.4 Aims and Scopes

This thesis aims to propose a novel framework that integrates Health IoT with HSNs to enable automated searches of relevant health information using biosignals collected from IoT sensors. The aims and scopes are organized according to the three main health conditions addressed: heart disease, respiratory disease, and bowel disease.

For heart disease, the first aim is to utilise biosignal analysis to help individuals with specific health conditions find relevant online support groups. The goal is to enable direct searching of online support groups using biosignals, improving the efficiency and user experience compared with manual searches. The second aim is to enable efficient retrieval of relevant information and expand the coverage of healthcare data possibilities beyond predefined diseases and abnormalities, specifically in the context of cardiovascular diseases and ECG signals. The third aim is to enable semantic search of text-based content in HSNs using ECG data from Health IoT. The goal is to create a human-readable format that describes abnormalities using measurable physiological measurements, enabling semantic searching of these descriptions within the HSNs, allowing direct comparison with textual data, and improving search capabilities.

In the case of respiratory disease, the first aim is to develop an efficient deep learning model that integrates the strengths of CNN, RNN and Transformer to achieve non-invasive diagnosis of respiratory diseases using breath sounds collected with mobile devices. The goal is to address the limitations of conventional deep learning methods in capturing diverse spatial relations and long sequences and to reduce overfitting. The second aim is to transform breath sounds into interpretable clinical measurements that represent specific physical abnormalities, such as breath frequency and depth. These clinical measurements will serve as keywords for semantic search within the HSNs, enhancing the search capabilities and facilitating targeted information retrieval.

Regarding bowel disease, the first aim is to develop cost-effective methods for collecting bowel sounds and extracting important information about bowel health conditions in the NICU to detect bowel diseases in newborns. The goal is to achieve real-time, non-invasive monitoring and diagnosis of gastrointestinal problems in newborn infants by automating the monitoring and detection of specific bowel movements. The second aim is to develop deep learning models that accurately segment and detect bowel sounds to improve the diagnosis of gastrointestinal problems, despite the challenges presented by the irregular pattern of bowel sounds. The third aim is to convert bowel sounds into interpretable clinical measurements that represent specific physical abnormalities analysed in Chapter 8. These measurements will include characteristics such as hypoactive bowel sounds and hyperactive bowel sounds and will serve as keywords for semantic searches within the HSNs. This conversion process will enhance the search capabilities of the HSN and facilitate targeted information retrieval.

1.5 Methodology

Figure 1.5 illustrates the overall process and architectures of the proposed integrated HSNs to achieve the aims. The methodology of this thesis is tailored to address the specific challenges related to heart disease, respiratory disease, and bowel disease, as identified in the aims and scopes.

In Chapter 3 Section 3.1, we propose integrating heart sound analysis into existing HSNs to enable remote diagnosis of heart diseases. We extract time series and Mel-Frequency Cepstral Coefficients (MFCC) features from the heart sounds and use the Gaussian Hamming Distance to compare healthy heart sounds with diseased heart sounds.

In Chapter 3 Section 3.2, we propose a Keyword-based Integrated HSN of Things (KIHoT). KIHoT utilises CNN to predict the word embeddings of medical keywords from ECG signals. These medical keywords are selected from HSN posts that are labelled as heart diseases associated with the corresponding ECG signals. The framework enables a direct comparison of ECG signals with textual data within HSNs, allowing for the exploration of healthcare data possibilities beyond predefined diseases and abnormalities.

In Chapter 3 Section 3.3, we have developed CardioVec, a comprehensive framework comprising four modules: the ECG wave detector, the ECG physiological measurement extractor, the ECG abnormality detector and the IoT-HSN search interface. The ECG wave detector identifies and locates specific positions of the ECG waveforms, from which physiological measurements are extracted. These measurements undergo further processing to generate human-readable keywords. Finally, we apply Sentence BERT (SBERT), modified BERT (Bidirectional Encoder Representations from Transformers) for Semantic Text Similarity (STS) (Reimers & Gurevych, 2019), to search for relevant information from the internet using the generated clinical descriptions as keywords.

In Chapter 4 Section 4.1, we propose a novel deep learning algorithm named Recurrent Local Relation Transformer Encoder Classifier (ReLATEC) that utilises relative positional encoding, multi-

head local attentions and recurrent encoder blocks to non-invasively diagnose respiratory diseases using sounds collected with mobile devices.

In Chapter 4 Section 4.2, we develop PhysioVec, a multistage deep-learning framework for indexing breath sounds and for directly searching HSNs using bowel sounds. The breath sound recordings are collected non-invasively using mobile phones. We extract breath clinical descriptions by segmenting the breath sound using an improved deep learning approach based on ReLATEC, LRT. We use the clinical descriptions to search online documents. We use SBERT to convert both online documents and clinical descriptions into vectors for semantic searching.

In Chapter 5 Section 5.1, we collected a total of 240 bowel sounds from two NICU hospitals using a miniature Bowel Sounds Sensor (BoSS). The 3D-printed micro stethoscope of BoSS enables continuous monitoring of bowel sounds in newborn infants with minimal discomfort. In this study, we annotated bowel sound events in newborn infants and performed clinical analysis of bowel sounds in term newborn infants. Based on the annotations, we identified physiological measurements of bowel sounds that can be used for diagnosing gastrointestinal health.

In Chapter 5 Section 5.2, we propose the segmentation of bowel sounds using ReLATEC. The locations of bowel sounds are then labelled to train the bowel sound detector. The labels are converted into weak labels with lower temporal resolution because it is difficult for human listeners to accurately identify all bowel sounds due to hospital environmental noises. We then train a ReLATEC model to detect the locations and two types of bowel sounds: short bursts and long bursts. The trained detector provides real-time annotation of bowel sounds with locations, types and statistics of bowel sound activities.

In Chapter 5 Section 5.3, we integrate bowel sounds into PhysioVec. We collect bowel sound recordings non-invasively using BoSS. We extract the clinical descriptions from the bowel sounds by segmenting it using LRT. We use the clinical descriptions to search online documents. We apply SBERT to convert both online documents and clinical descriptions into vectors for semantic searching.

1.6 Contributions

The major contribution of this thesis is the integration of HSNs and IoT by interpreting biosignals to be searchable in the HSNs that solves the problems illustrated in Figures 1.2 and 1.3. We first perform a comprehensive survey on current existing HSNs and IoT to identify that there are no current existing methods for integration of HSNs and IoT.

The contribution of Chapter 3 Section 3.1 is the introduction of Gaussian Hamming Distance as a method to identify similar heart condition groups within HSNs. By comparing heart sounds among users within the HSNs, this approach enables the automatic matching of patients with similar conditions based on their heart sounds. This integration of heart sound analysis into HSNs improves the accuracy and convenience of remote diagnosis for cardiovascular disease, eliminating the need for

patients to physically visit medical facilities for specialised testing. Our publication (Yi Huang & Song, 2018) contributes to this chapter.

The contribution of Chapter 3 Section 3.2 is the development of the Keyword-based Integrated HSN of Things (KIHoT), which predicts keywords from ECG signals and utilises them to search the HSNs for content with similar keywords. This integration of ECG signals and health social networks improves the prediction of heart conditions and facilitates comprehensive access to relevant information. Our publication (Y. Huang & Song, 2019) contributes to this chapter.

The contributions of Chapter 3 Section 3.3 include providing a feasible solution for HSN semantical searching by translating biosignals into human-readable keywords. The integration of IoT and HSNs allows doctors and patients to search relevant medical information directly using ECG biosignals. Our keyword generation is evidence based and is more readable and more informative for doctors and patients. Our publication (Y. Huang & Song, 2022a) contributes to this chapter.

The contribution of Chapter 4 Section 4.1 is introducing ReLATEC as a novel deep learning algorithm specifically designed for non-invasive diagnosis of respiratory diseases using breath sounds collected with mobile devices. The algorithm combines the advantages of CNN, RNN and Transformer models, leveraging their strengths in feature composition, temporal modelling and flexible spatial relations. The evaluation of ReLATEC against CRNN and vanilla Transformer demonstrates its superior performance and reduced overfitting, highlighting its potential for accurate and efficient respiratory disease diagnosis.

The contributions of Chapter 4 Section 4.2 are: (1) We proposed a novel LRT combining local attention and recurrent Transformer to reduce overfitting and improve performance in the segmentation of breathing sounds. (2) Interpret breath pattern abnormality by extracting physiological measurements. (3) PhysioVect successfully integrated IoT with HSNs allowing patients and doctors to directly search relevant health information using breathing sounds. Our publication (Y. Huang & Song, 2022b) contributes to this chapter.

The contributions of Chapter 5 Section 5.1 are: (1) This is the first time we have recorded bowel sounds of newborn infants during pre-feed and post-feed using miniature IoT sensors continuously. (2) We measure five clinical bowel activities and classify their types into short bursts (ticks) and long bursts (grumbles). (3) We use an arithmetic median (median without duplicates) for filtering the physiological measurements of bowel sound activities for hospital environments and sensor misplacement and alignments. (4) We find five physiological measurements that are significantly different between sick and healthy infants: tick rate, tick duration, tick power, grumble duration and grumble power.

The contribution in Chapter 5 Section 5.2 is the collection of bowel sounds from newborn infants using BoSS and the utilisation of this data to train the ReLATEC deep learning model with novel weak annotations. This contribution enhances the training process of the ReLATEC model by

incorporating relevant bowel sound data from newborn infants. Our publication (Song, Huang, Koh, & Rajadurai, 2021) contributes to this chapter.

The contributions of Chapter 5 Section 5.3 are: (1) Detection and segmentation of bowel sounds into two different types of bowel activities using a Local Recurrent Transformer (LRT) that combines local attention and recurrent transformer to reduce overfitting and improve the performance of bowel sound segmentation; (2) Extracting human interpretable physiological information through intestinal sound pattern; (3) Automatically searching relevant physiological information of bowel activities with extracted physiological information on the internet. Our publication (Y. Huang & Song, 2022c) contributes to this chapter.

Due to the time limit, only the final method and important intermediate method includes ablation study, and we leave the ablation studies the other preliminary and intermediate studies as future work.

2 Literature Review

This chapter presents a literature review that focuses on recent research related to the application of deep learning in Health IoT and HSNs. The review encompasses deep learning of biosignal analysis, Health IoT, and information retrieval in HSNs. The primary objectives of this literature review are to identify unexplored or unsolved problems and to determine the state-of-the-art approaches.

To conduct this literature review, we collected relevant studies from search results and references of other papers. Each article was thoroughly reviewed, analysing, and summarising the topics, problems, methodologies, and results. The search involved exploring academic publication databases to identify papers published within the past five years. These papers were then reviewed to analyse the addressed topics, problems, methodologies and reported results. A meta-analysis was conducted to synthesise the findings, and identify trends, opportunities, and gaps in the literature. Specifically, we conducted a systemic review of recent relevant research on deep learning for signal processing, Health IoT and HSNs. We also reviewed the technical backgrounds of deep learning and compared the advantages and disadvantages of current approaches.

For the collected papers' applications and methodologies, we calculated the weighted average accuracy. This measurement allowed us to compare the weighted accuracy, highest accuracy and lowest accuracy across different application areas, such as bowel activity analysis, lung sound analysis and heart sound analysis. Additionally, we compared and presented the weighted average accuracy of different data mining algorithms to evaluate the performances of various approaches. The weighted average accuracy serves as a performance metric for identifying the best technologies and identifying gaps in certain approaches. The calculation of the weighted average accuracy is as (2.1)

$$A = \frac{\sum a_i}{\sum m_i}, \tag{2.1}$$

a_i = Overall accuracy of research i

m_i = Number of samples in the research i

Through our literature review, we identified a trend, research gap and opportunity. The trend observed is that HSNs and IoT are becoming important sources of health-related information. However, there is currently no integration between Health IoT and HSNs. This research gap presents an opportunity for the development of deep learning as a new approach for searching online information and processing biosignals.

In summary, this literature review examines recent research on the application of deep learning in Health IoT and HSNs. The Health IoT section focuses on automated biosignal analysis and the diagnosis of diseases such as heart conditions, respiratory illnesses and gastrointestinal issues. The

aim is to find reliable automated diagnosis methods using biosignals, addressing the need for objective, non-invasive and low-cost diagnostic techniques. Common methods employed include CNN, RNN and other deep-learning approaches that analyse sounds such as heart murmurs, wheezes and bowel noises. The key contribution of these studies is demonstrating that deep learning can provide accurate automated diagnosis.

In the HSN mining section, the main topics revolve around detecting mental illnesses, adverse drug reactions and influenza spread using social media data. The challenges in this domain include a lack of data in conventional studies and difficulties in interpreting informal texts. The aim is to access large volumes of data for healthcare insights. Common methods utilised are Support Vector Machine (SVM), RNN and CNN, which analyse the text and metadata of posts. The contributions of these studies include showcasing the value of social media for public health monitoring and overcoming the limitations of traditional data collection methods.

The main contributions of this literature review are: 1) Providing an overview of the current applications, approaches and performances in applying deep learning to biosignals, Health IoT and HSNs. 2) Identifying open research questions and problems in integrating these domains. 3) Suggesting potential opportunities for future work in developing artificial intelligence solutions for healthcare using multimodal data sources.

2.1 Health IoT

This section discusses the application areas, current problems, and approaches of Health IoT for the automated diagnosis of biosignals, including heart sounds, ECG, respiratory sounds and bowel sounds. Health IoT enables doctors to remotely monitor patients, and advancements in signal processing and machine learning have facilitated automated diagnosis for diseases related to the bowel, lungs, and heart.

2.1.1 Application Areas

Heart sounds are widely used in the diagnosis of coronary artery disease and heart valve defects. Abnormal heart sounds can include additional components, such as extra heart sounds or murmurs. The detection of abnormal heartbeats indicates underlying diseases, including dysrhythmia, valvular disease, stenosis, or pericarditis (F. Wang, Syeda-Mahmood, & Beymer, 2007). Heart sound classification methods have been used for diagnosing heart valve defects and other heart diseases (Bozkurt, Germanakis, & Stylianou, 2018; Kucharski, Grochala, Kajor, & Kantoch, 2018; Nilanon, Yao, Hao, Purushotham, & Liu, 2016; Potes, Parvaneh, Rahman, & Conroy, 2016; Rubin et al., 2016). Murmurs are abnormal heart sounds caused by turbulent blood flow through a heart valve and can be either innocent, indicating a harmless condition, or indicative of a serious heart condition (Dominguez-Morales, Jimenez-Fernandez, Dominguez-Morales, & Jimenez-Moreno, 2018). Several

studies have specifically focused on the detection of murmurs in heart sounds (Deng & Han, 2016; Deperlioglu, 2018; Dominguez-Morales et al., 2018; W. J. Zhang, Han, & Ieee, 2017).

ECG, another type of biosignal, is used in addition to heart sound signals for diagnosis (Yildirim, 2018). ECG provides information about heart function and conditions, such as heart arrhythmia. Classifiers such as SVM (Azariadi, Tsoutsouras, Xydis, & Soudris, 2016; Verma & Sood, 2018), CNN (Xiong et al., 2018), and RNN (Yildirim, 2018) have been used with ECG signals. Some approaches have been applied to IoT heart diagnosis systems using ECG signals (Azariadi et al., 2016; Verma & Sood, 2018).

Bowel sounds are non-stationary audio signals produced by the gastrointestinal system that may carry physiological information about gastrointestinal health. Previous studies have focused on non-invasive monitoring of intestinal sounds by analysing jitter and shimmer to extract physiological information (K. S. Kim, Seo, Ryu, Kim, & Song, 2011). Some studies have compared the similarity of bowel sounds to investigate their characteristics (Y. Huang et al., 2017). Some studies have focused on the long-term monitoring of gastrointestinal motility by detecting and analysing bowel sounds (Goto et al., 2015; Y. Huang & Song, 2019; Yin et al., 2018). Automated diagnosis based on biosignals can be applied to bowel diseases such as irritable bowel syndrome and the detection of abnormal bowel sounds (Dimoulas, Kalliris, Papanikolaou, Petridis, & Kalampakas, 2008; Y. Huang et al., 2017). Bowel sound detection is also used to measure the frequency of bowel sounds in various studies (Sazonov et al., 2010; Yin et al., 2018).

Respiratory sounds, also known as lung sounds or breath sounds, are non-periodic and non-stationary sounds produced by the lungs and respiratory system (Marshall & Boussakta, 2007). These sounds contain important indicators of respiratory diseases such as flu, pneumonia, bronchitis, chronic obstructive pulmonary disease, and the severity of bronchial airflow limitation. Breath sounds can be collected non-invasively using a handheld microphone placed near the mouth and nose (B. Lei, Rahman, & Song, 2014; I. Song, 2015) aiding in the diagnosis of conditions such as pneumonia and other respiratory diseases. There are two types of respiratory sounds: normal breathing sounds and abnormal sounds. The normal breath process produces breath sounds, while adventitious sounds are usually related to abnormal mechanisms. Abnormal respiratory sounds can be further divided into two categories: continuous sounds and discontinuous sounds (Bahoura, 2009; Charleston-Villalobos et al., 2011). Wheezing, a whistling or squeaky continuous sound, is a common indicator of asthma, especially in children (Bahoura, 2009; Bokov, Mahut, Flaud, & Delclaux, 2016; Kochetov, Putin, Azizov, Skorobogatov, & Filchenkov, 2017). Some studies focus on detecting wheezing in breath sounds specifically, while others focus on detecting various types of abnormal respiratory sounds, including rhonchi and crackles (Aykanat, Kilic, Kurt, & Saryal, 2017; Bardou, Zhang, & Ahmad, 2018; İçer & Gengeç, 2014; Naves, Barbosa, & Ferreira, 2016; Sengupta, Sahidullah, & Saha, 2016).

2.1.2 Problems

2.1.2.1 The Need for an Objective Diagnosis Method

Manual auscultation is a subjective process and not always able to provide accurate results (Aykanat et al., 2017; Bahoura, 2009; Deperlioglu, 2018; Dominguez-Morales et al., 2018; Y. Huang et al., 2017; Naves et al., 2016; Yin et al., 2018), which can be affected by factors such as the use of a stethoscope, the low sensitivity of the human ear to low-frequency sounds and the subjective experience of trained professionals in traditional heart auscultation (Charleston-Villalobos et al., 2011; Deng & Han, 2016; Sengupta et al., 2016). In the case of wheeze recognition, parents of children are only able to identify wheezing in 59% of cases (Bokov et al., 2016). Indirect bowel disorders diagnosed with symptoms can be strongly affected by different unrelated factors such as nutrition, stress, and drug usage (Dimoulas et al., 2008). Bowel sound auscultation is proposed for a more objective diagnosis. Additionally, the use of handcrafted features and conventional classification approaches can introduce subjectivity, whereas deep learning enables end-to-end model building, reducing the reliance on subjective interpretation (Yildirim, 2018).

2.1.2.2 The Need for Non-Invasive Diagnosis

Bowel sound analysis requires minimal physical contact with the patient's skin surface, making it a more comfortable and non-invasive alternative that does not require specialised professionals (Y. Huang et al., 2017). Manometry and electromyography, which are invasive, can cause discomfort and inconvenience (Dimoulas et al., 2008). Bowel sound analysis only needs minimal physical contact to the skin surface of patients, so it is more comfortable and does not need to be operated by trained professionals. However, the author proposes a sound-based method for detecting swallowing events to monitor food intake.

Most computerised breath sound analysis methods rely on contact-based data collection, which not only discomforts patients but also requires specialised medical devices and trained professionals (B. Lei et al., 2014; I. Song, 2015). Advanced techniques such as X-ray, spirogram and arterial blood analysis are invasive, expensive, and potentially harmful (Sengupta et al., 2016).

2.1.2.3 Environmental Noise

Environmental noise poses a challenge in non-invasive computerised breath sound analysis (I. Song, 2015) and the long-term monitoring of bowel activities (Dimoulas et al., 2008), as contactless methods introduce environmental noise. One possible countermeasure is to record the breath sound with the left channel and subtract it from the environmental noise recorded by the right channel to eliminate noise. Accurate and robust approaches are needed for processing low-quality signals, including segmenting heart sounds from murmurs and other noise (Deng & Han, 2016; Kang et al., 2018; Rubin et al., 2016).

2.1.2.4 The Need for a Low-Cost Diagnosis Method

Advanced techniques such as X-ray, spirogram and arterial blood analysis are invasive, which can be expensive and potentially harmful (Griffel, Zia, Fridman, Saponieri, & Semmlow, 2013; Sengupta et al., 2016). Additionally, complex technologies such as echocardiography, computerised tomography and magnetic resonance imaging require specialised expertise for operation.

2.1.3 Methods

The conventional approaches to biosignal-based diagnosis involve several steps: pre-processing, segmentation, feature extraction and classification. In the pre-processing step, the sounds are resampled and normalised. A segmentation algorithm is typically used to identify the portions of the biosignal that contain relevant information, such as heartbeats or bowel activities. Researchers often employ specific feature extraction methods, such as Fourier transformation, wavelets or derivatives, to summarise the frequency information and other relevant features of the signal. These extraction methods require expert knowledge and manual work. Fortunately, deep learning methods offer end-to-end approaches that can accept raw ECG signals or audio spectrograms as input and generate features based on the desired tasks. Some deep learning approaches do not require segmentation, as they can automatically locate relevant information and discard irrelevant information from raw signals (Xiong et al., 2018). Segmentation plays a crucial role in identifying the presence of specific sounds and patterns, enabling the detection of abnormalities. For instance, segmentation can detect exhalations during breath sound periods, helping identify irregularities. Classification, on the other hand, is a machine learning concept where different algorithms are trained using datasets that contain relevant features. Once trained, these algorithms can classify new sounds based on their learned patterns. Popular algorithms for heart sound classification include deep learning and conventional machine learning techniques.

2.1.3.1 Features

The common features of audio biosignals are MFCC, Wavelet Transform, Fourier Transform, Short-Time Fourier Transform (STFT), and other features. Fourier transform-based features describe the frequency characteristics of the signal. Fast Fourier Transform (FFT) are widely used in different types of audio (Amit, Gavriely, & Intrator, 2009; Verma & Sood, 2018; Yin et al., 2018).

MFCC is widely used in sound signal analysis (Bahoura, 2009; B. Lei et al., 2014; Sengupta et al., 2016). MFCC can be estimated by using a parametric approach derived from Linear Predictive Coding (LPC), or by using a nonparametric FFT-based approach (B. Lei et al., 2014). MFCC has been used for breath sound analysis and bowel sound analysis (Y. Huang et al., 2017; B. Lei et al., 2014; Sazonov et al., 2010; I. Song, 2015).

There are other features such as Power Spectral Density (PSD) (Bokov et al., 2016; Charleston-Villalobos et al., 2011; İçer & Gengeç, 2014; I. Song, 2015), Autoregressive (Charleston-Villalobos et

al., 2011), LPC (Bahoura, 2009), time domain (Amit et al., 2009; Dimoulas et al., 2008; Y. Huang et al., 2017).

2.1.3.2 Convolutional Neural Network (CNN)

CNN is a type of feed-forward neural network that draws inspiration from the biological concept of visual processing and was initially designed for interpreting two-dimensional images. It builds on the idea of relating signals from a fixed relative distance to learn local dependencies. In a CNN, a convolutional layer consists of a set of feature maps. Each feature map uses a set of trainable shared weights to detect elementary features from the previous layer using a sliding window. Positional accuracy in a feature map is less critical and can be detrimental when there is a shift in the input. To down sample the feature map and reduce sensitivity to the exact position, a pooling layer is often used, which selects the maximum value within a region. Compared with a conventional, fully connected neural network, CNN requires less memory and exhibits greater robustness to shift, scale and distortion. The activation function commonly used in CNNs is the ReLU (Restricted Linear Unit) function.

Equation (5.16) is an expression of a convolution layer.

$$(f * g)(x, y) = \sum_{s-a}^{a} \sum_{t-b}^{b} f(s, t) g(x - s, y - t), \tag{2.2}$$

where $f(x, y)$ is the input image, which in a CNN usually represents the pixel values of a small region of the image, such as a 5x5 patch. $g(s, t)$ is the convolution kernel (or filter), which is a small matrix of weights, typically smaller than the input image, such as 3x3 or 5x5. $*$ denotes the convolution operation. (x, y) is a pixel in the output image. s ,t are the positions of elements in the convolution kernel. a, b are the half-width and half-height of the convolution kernel.

Instead of the sigmoid function, CNN usually uses the *ReLU* (Restricted Linear Unit) function as its activation function as in (2.3).

$$ReLU(x) = max\ (0, x), \tag{2.3}$$

where X is the training loss; W and b are trainable weight and bias; m and n are the height and width of the sliding window.

One limitation of CNN arises from its fixed distribution of feature maps, whether spatial or temporal, which can hinder the learning of meaningful combinations of features. This limitation makes CNN inefficient in modelling varying spatial relationships (Vaswani et al., 2017). Conventional approaches to address this problem involve using more filters and data to learn different combinations of features in different positions. However, when the available data is insufficient, the performance of the classifier is significantly compromised.

2.1.3.3 Recurrent Neural Network (RNN)

RNN is another type of neural network that is highly preferred for processing sequential data (Yildirim, 2018). In an RNN, the outputs of the neurons are recursively fed back to their inputs, allowing them to perform the same task for every input. Long Short-Term Memory (LSTM) is the most used type of RNN. It utilises a structure known as a memory block, which consists of three gates: the input gate, the output gate and the forgetting gate. Each gate functions as a simple neural network, enabling LSTM to effectively store and retrieve information over time.

RNN has been widely used in Natural Language Processing (NLP) and other time sequence analyses due to its ability to relate each signal recursively (Dehghani, Gouws, Vinyals, Uszkoreit, & Kaiser, 2018; Vaswani et al., 2017). However, conventional RNNs have limitations when handling long sequences. As the gradient accumulates along the input sequence during training, a long sequence can cause the gradient to either vanish or explode. This phenomenon can significantly impact the ability of the RNN to learn and capture long-range relationships in the data.

2.1.3.4 Word Embedding

Word embedding is a widely used method in NLP that projects words into a vector space. Unlike one-hot encoding, word embedding captures important semantic features of words. For example, similar words like “cough” and “breathless” have similar values in word embedding, indicating their semantic similarity. In contrast, their one-hot representations lack this similarity. One of the key features of word embedding is that it maintains similar vector distances between word pairs with similar relationships. Equation (2.4) shows an example of this property.

$$V_{king} - V_{man} \approx V_{queen} - V_{woman}, \tag{2.4}$$

where V means the embedding of a word.

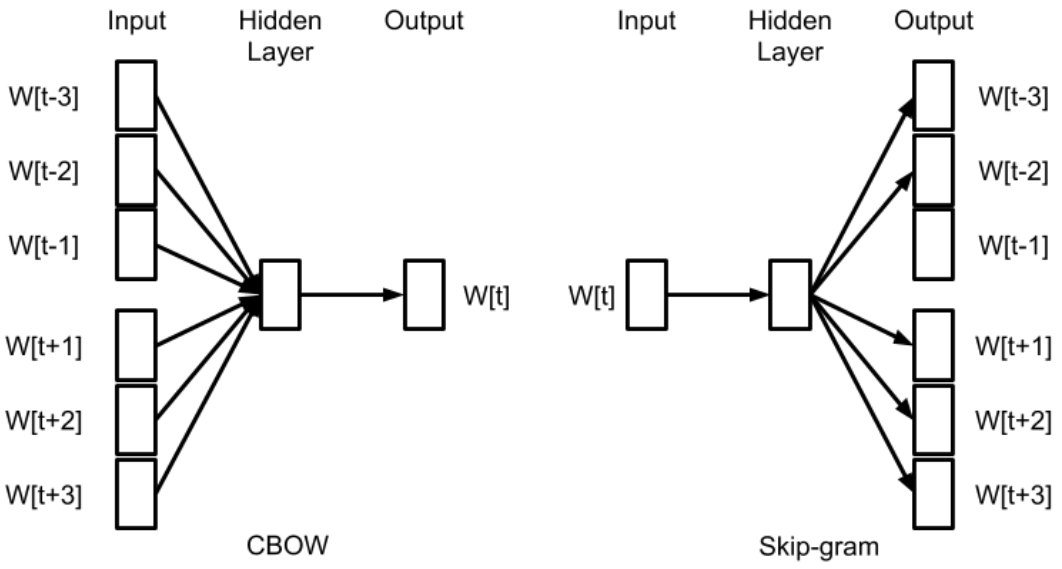

Figure 2.1 ***Illustration of CBOW and Skip-gram.***

Word embedding is generated through a feature learning approach, where word representations are learned from the input layer of a neural network. Two common word embedding models are Continuous Bag of Words (CBOW) and Skip-gram. In the CBOW model, a vector is generated by predicting the distribution of possible words using the context of the five words before and after the target word. In contrast, the Skip-gram model predicts the distribution of the context words given the input word. The learned word representations are then obtained from the weights of the hidden layer in the neural network. These vector representations can be stored and used like a dictionary, allowing for efficient retrieval and usage of word embeddings. Figure 2.1 provides a visual representation illustrating the differences between CBOW and Skip-gram models.

2.1.3.5 Attention and Transformer

The attention-based approach has been introduced to address the challenge of capturing long-range relationships in previous deep-learning approaches in both computer vision (CV) and NLP studies. One of the best-known attention-based approaches is the Transformer, initially designed for machine translation tasks (Vaswani et al., 2017). The input to the Transformer consists of word tokens annotated in a one-hot encoding, which is then processed as a sequence. The Transformer architecture has been adapted for CV and audio tasks, where word tokens are replaced by image features and audio features such as MFCC and spectrogram.

Unlike CNN and RNN, the Transformer models relations between time steps using an attention mechanism. Unlike RNN, which has a varying number of operations depending on the input sequence length, the Transformer has a fixed number of operations, resulting in more efficient processing. Additionally, the Transformer learns relations without relying on fixed positions, in contrast to CNN (Hu et al., 2019). The attention mechanism in Transformer allows the model to capture relationships based on dot-product similarity among input time steps, enabling dynamic learning of relations. This contrasts with CNN, which is limited in capturing relations within fixed distances.

In Transformer, Scaled Dot-Product Attention is used in (2.6):

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{2.6}$$

where Q, K and V are query, key and value, respectively; and d_k is the dimension of the key. Attention function maps query vectors and key-value vectors by dot-product similarity. The attention mechanism allows the modelling of input and output relations independent of their positions (Vaswani et al., 2017).

Based on attention, Transformer uses multi-head attention *MHA* to learn different types of relations as shown below:

$$MHA(Q,K,V) = concat(head_0, \dots, head_h)W_o, \quad (2.7)$$

$$head_i = Attention(QW_i^q, KW_i^k, VW_i^v), \quad (2.7)$$

where W denotes learnable weights; Q, K, V denote query, key, value; W^q, W^k, W^v denotes weights of query, key, value; h denotes the total number of heads; i denotes the index of the head. The concatenated vector is then multiplied by a parameter W_o.

2.1.3.6 Local Relation Modelling Based on Local Attention

The self-attention feature in the Transformer model calculates attention between all input nodes, enabling the capture of global relations. In contrast, local relation approaches calculate attention within a specific aggregation window, capturing local attention energy among inputs. This aggregation window introduces an information bottleneck like CNN, which is crucial for feature learning in the field of CV (Hu et al., 2019).

One example of a local relation approach is the Local Relation Network (LR-Net) (Hu et al., 2019), which focuses on finding pixel-to-pixel relations in images. LR-Net replaces the convolutional layers in ResNet with local relation layers, enabling the extraction of local relations within the image.

Building upon LR-Net, the Swin Transformer (Liu et al., 2021) has been proposed and has achieved state-of-the-art performance in image classification tasks. The Swin Transformer further enhances the capabilities of LR-Net, although the specific improvements and mechanisms that contribute to its superior performance require further elaboration.

2.1.3.7 Recurrent Transformers

Recurrent Transformers refer to a variant of Transformers that incorporate a recurrent encoder block. Examples of Recurrent Transformers include ALBERT (Lan et al., 2019) and Universal Transformer (UT) (Dehghani et al., 2018). In this context, a recurrent encoder block consists of an attention layer and a feed-forward non-linear layer.

Instead of using multiple encoder blocks with independent learnable parameters, the recurrent encoder block in Recurrent Transformers iteratively computes contextual features. This means that the recurrent encoder block is responsible for computing contextual features across multiple layers, utilising shared parameters. This approach, known as cross-layer parameter sharing, allows for efficient computation and captures long-range dependencies within the input sequence.

The addition of a recurrent structure in Recurrent Transformers brings several advantages. Firstly, it leads to a significant reduction in parameters through cross-layer weight sharing. This weight sharing allows for efficient parameter usage across layers. Secondly, the recurrent structure introduces recurrent prior knowledge like RNNs, which enables the model to capture sequential dependencies and temporal information more effectively (Dehghani et al., 2018).

One example of a model that incorporates a recurrent encoder block is UT (Dehghani et al., 2018). UT was proposed as an extension to the vanilla Transformer, which is not Turing complete and cannot perform certain tasks that can be easily handled by RNNs. By introducing the recurrent encoder block, UT is Turing complete, meaning it can theoretically compute any computable function given sufficient resources and time. This property of being Turing complete opens a broader range of tasks that UT can handle effectively.

In contrast to UT, ALBERT (Lan et al., 2019) incorporates cross-layer parameter sharing to reduce the number of parameters and has demonstrated superior performance compared with approaches that use independent parameters. One significant difference between ALBERT and UT lies in their treatment of iterations. ALBERT employs a fixed number of iterations (12 times), while UT dynamically halts the iterations based on certain criteria. The success of ALBERT suggests that a fixed number of iterations of a single Transformer Encoder block can yield satisfactory performance.

2.1.3.8 Common Approaches to Analysing Diagnostic Sounds

Table 2.1 ***Feature matrix of common approaches to analysing diagnostic sounds.***

Approaches	Method (Features)	Accuracy	Data Size	Reference
CNN	MFCC	84.0%	3240	(Rubin et al., 2016)
	DWT	82%	3240	(Potes et al., 2016)
		85%	3240	
	MFCC	81.3%	3240	(Nilanon et al., 2016)
	MFCC	81.4%	3240	(Bozkurt et al., 2018)
	Fourier Transform	95.2%	3240	(Kucharski et al., 2018)
	RGB image	94.2%	3240	(Deperlioglu, 2018)
	Sonogram	94.2%	3126	(Dominguez-Morales et al., 2018)
	Spectrogram	89.8%	50	(Kang et al., 2018)
	Spectrogram	86%	1630	(Aykanat et al., 2017)
	Spectrogram	98.5%	50	(Bardou et al., 2018)
	STFT Spectrogram	98.2%	817	(Kochetov et al., 2017)
	Spectrogram	74.0%	176	(W. J. Zhang et al., 2017)
	Mel Spectrogram	57.91% Macro F1-score, 66.38%	920	Classification of wheeze and crackle

		Micro F1-Score		(Acharya & Basu, 2020)
	Mel Spectrogram	89.52% F1-score	543	COVID-19 detection (Imran et al., 2020)
	MFCC	0.95 AUC (Area under Curve)	2660	COVID-19 detection (Laguarta, Hueto, & Subirana, 2020)
	ECG	90.6% Score	8528	Atrial Fibrillation Detection (Limam, Precioso, & Ieee, 2017)
	ECG Time Domain	86.4%	12186	(Xiong et al., 2018)
Weighted Average		87.0%		
RNN		89.9%	8528	(Limam et al., 2017)
	ECG	99.39%	360	(Yildirim, 2018)
Weighted Average		90.3%		

In Table 2.1, we can find out it is reliable to diagnose disease through sound analysis with classification. CNN has been used in the diagnosis of COVID-19 with cough sound (Imran et al., 2020; Laguarta et al., 2020). Acharya and Basu (2020) proposed respiratory sound classification for normal, wheeze, crackle, and both (wheeze and crackle) sounds using a CRNN based on LSTM and CNN. LSTM uses a memory block structure to build a long-distance relations network, which is the most common RNN approach. Limam et al. (2017) also previously proposed a phonocardiogram classification method (containing both ECG and heart sound) in the 2016 PhysioNet challenge dataset using CRNN. Yildirim (2018) proposed ECG classification with BiLSTM (Bidirectional LSTM) and Discrete Wavelet Transform (DWT) features. The common procedure for respiratory sound diagnosis using deep learning is to extract frame features from the audio and then send those features to the model.

Automated diagnosis offers objectivity and accuracy compared with manual auscultation. Health IoT addresses the need for non-invasive, objective, pervasive and cost-effective diagnostic methods. It overcomes the limitations of subjective manual diagnosis and has been proven reliable in the literature. Currently, Health IoT demonstrates potential in various medical applications, particularly in disease diagnosis and prediction. For instance, studies have focused on early detection of heart disease (Kumar & Gandhi, 2018; Verma & Sood, 2018) and monitoring (Azariadi et al., 2016; Manogaran et al., 2018).

In summary, Health IoT, coupled with signal processing and machine learning techniques, provides opportunities for automated diagnosis of biosignals. It offers objective and accurate diagnostic methods for diseases related to the bowel, lungs, and heart, overcoming the limitations of subjective manual diagnosis. Literature supports the reliability and potential of Health IoT in various medical applications, including disease detection and monitoring.

2.2 HSN Mining

In this section, we review and analyse relevant studies in the field of HSN mining, focusing on their application areas, specific problems addressed, and methods employed. The healthcare sector has undergone significant transformations in recent years, largely due to the increasing connectivity of the internet. This connectivity has resulted in a growing reliance on the internet for making healthcare decisions. A healthcare survey conducted in 2011 revealed that 80% of internet users have sought online information related to healthcare (Fox, 2011). Additionally, the survey found that one in five adult internet users have utilised online platforms to connect with others who share similar health concerns (Fox, 2011). These findings highlight the importance of exploring HSN mining and its implications for online health communities.

HSNs are online communities driven by patients that play a crucial role in providing valuable information to patients and caregivers (Song & Marsh, 2012). While medical professionals and offline resources form the foundation of healthcare, the need for knowledge-sharing on social networks remains significant. HSNs have emerged as an important source of information for both patients and physicians, enabling them to make well-informed decisions regarding healthcare. Through platforms like social media, millions of users collaborate to generate a vast pool of medical knowledge. For instance, a study found that 20% of adults have discovered individuals with similar health problems online, and most of them have benefited from HSNs in various ways (Fox, 2011). HSNs also facilitate connections among individuals who share the same health conditions, enabling them to exchange information and provide emotional support.

Various online platforms are available for healthcare purposes. Blogs, which are updatable journals/diaries published by users, offer a means for self-publishing information and managing web links. However, the quality of information in blogs cannot always be guaranteed (Miller & Pole, 2010). Blogs also serve as a platform for communication between medical experts and patients, fostering open dialogue (Kamel Boulos & Wheeler, 2007). Wikis, on the other hand, are collaborative websites where users can update and contribute content. Although popular and user-friendly, wikis are susceptible to vandalism. One notable example is the Flu Wiki, which effectively raises awareness of influenza within local public health communities (Boulos, Maramba, & Wheeler, 2006). Additionally, there are social sites that provide users with a vast database of healthcare providers to choose from. Websites like “Patient Option” enable patients to provide feedback and rate doctors online, allowing other patients to make informed decisions based on these ratings (Kamel Boulos & Wheeler, 2007).

Adverse medical reactions, unfortunately, have resulted in unnecessary expenditures of billions of dollars.

In this thesis, HSNs encompass a range of platforms, including general-purpose social networks like Twitter (now *X*), that engage in discussions related to health issues and offer a wealth of health-related data (Achrekar, Gandhe, Lazarus, Yu, & Liu, 2013; Alimova & Tutubalina, 2017; Coppersmith, Dredze, & Harman, 2014; Miftahutdinov et al., 2017; Na et al., 2012; Sharif, Zaffar, Abbasi, & Zimbra, 2014; Sun, Ye, & Ren, 2015; Elena Tutubalina, Miftahutdinov, Nikolenko, & Malykh, 2018; Elena Tutubalina & Nikolenko, 2017). This thesis also considers HSNs to encompass various sources of health-related information, such as online health forums, government health statistics, medical reports, journal articles, drug databases and other medical information sources.

HSNs have been proposed by researchers as a cost-effective solution for enabling patients to support each other by sharing their experiences online (Song & Marsh, 2012). In addition, various automatic recommendation systems have been developed for social networks (Song & Marsh, 2012). Compared with conventional methods of data collection, collecting data from HSNs is faster, more cost-effective and yields larger amounts of data. Current approaches in the field primarily focus on retrieving information from HSNs related to mental health (Coppersmith et al., 2014; Song & Marsh, 2012; Elena Tutubalina & Nikolenko, 2017), the spread of influenza (Achrekar et al., 2013; Sun et al., 2015), and Adverse Drug Reactions (ADRs) (Alimova & Tutubalina, 2017; Miftahutdinov et al., 2017; Na et al., 2012; Sharif et al., 2014; Elena Tutubalina et al., 2018; Elena Tutubalina & Nikolenko).

2.2.1 Application Areas

Numerous studies have focused on detecting mental disorders within social networks, including the assessment of suicide risk and emotional disorders (Cheng, Li, Kwok, Zhu, & Yip, 2017). Furthermore, classification methods have been developed to identify depression, bipolar disorder, Post Traumatic Stress Disorder (PTSD), and seasonal affective disorder within *X*/Twitter (Coppersmith et al., 2014). An anonymous health condition indexing method has been proposed for conditions such as autism and Attention Deficit Hyperactivity Disorder (ADHD) (Song & Marsh, 2012).

In the realm of online cancer communities, research has examined the relationships between users and the identification of users (X. Wang, Zhao, & Street, 2014, 2017). Additionally, efforts have been made to identify influential users within HSNs for cancer survivors or caregivers (Zhao et al., 2014), explore the topics of discussion within breast cancer communities (S. Zhang, Grave, Sklar, & Elhadad, 2017), and predict whether cancer-related accounts are personal or organisational to analyse responses to chemotherapy side effects (L. Zhang, Hall, & Bastola, 2018).

ADRs are serious conditions caused by medications and play a crucial role in drug post-marketing (Elena Tutubalina & Nikolenko, 2017). Social media analysis is widely used for detecting ADRs due

to the availability of large amounts of data (Alimova & Tutubalina, 2017). Various approaches have been proposed for extracting ADR-related phrases in HSNs, including the use of SVM (Alimova & Tutubalina, 2017; Sharif et al., 2014), linguistic approaches (Na et al., 2012), and RNN (Elena Tutubalina et al., 2018; Elena Tutubalina & Nikolenko, 2017), as well as Conditional Random Fields (CRF) (Miftahutdinov et al., 2017; Elena Tutubalina & Nikolenko, 2017). In addition, Tutubalina et al. (2018) proposed a CNN approach for gender and age prediction in various health conditions and ADRs, achieving more than 65% F-measure for gender prediction and over 50% F-measure for age prediction.

The detection of the spread of influenza in social media offers a fast and cost-effective alternative to conventional approaches. Standard approaches have achieved F-measures of 0.814 on *X*/Twitter, 0.768 on Facebook (Achrekar et al., 2013), and 0.892 on Weibo, a Chinese social media platform (Sun et al., 2015).

2.2.2 Problems

2.2.2.1 Cost of Healthcare Service

Computerised medical services offer the potential to reduce healthcare costs. These services can be broadly categorised into four categories: automated diagnosis and assessment, online health support, telehealth, and health information management systems (Vong & Song, 2015). The goal of inclusive medical care is to provide affordable healthcare without compromising quality. To achieve this goal, healthcare services should be automated and available around the clock. Additionally, enabling remote payment methods such as mobile banking can further enhance accessibility and convenience (Vong & Song, 2015).

2.2.2.2 Finding Health-Related Information Online

The abundance of information available online poses challenges in the healthcare domain. While the vast availability of medical information empowers patients and encourages active involvement in their treatment discussions, it can also lead to difficulties in concluding personal health status. The presence of irrelevant information online hampers the process of accurately assessing one's health. Furthermore, healthcare information obtained online may not always align with the user's specific medical condition, increasing the risk of misinterpretation. The presence of medical jargon further complicates matters, making it challenging for individuals without medical expertise to understand the information they come across. Additionally, medical information necessary for patient-oriented decision-making is often scattered across various websites, making it difficult for individuals to access comprehensive and reliable information in one place (Wiesner & Pfeifer, 2014).

2.2.2.3 Absence of Data and Information

The major problem for data mining in HSNs to solve is the lack of data in conventional studies. Data mining in HSNs solves this problem because there is a large amount of data in the social network. For

mental health research, the traditional method is expensive, time-consuming (Coppersmith et al., 2014) and only has a little data (Elena Tutubalina & Nikolenko, 2017). In conventional approaches, inexperienced graduate students accessed suicide risk in social networks and boosted the proportion of suicidal cases (Cheng et al., 2017). For flu monitoring, the traditional method is by collecting data from doctors manually, which has a one- to two-week delay (Achrekar et al., 2013). It is also impossible to manually process a large number of texts on social media, and clinical trials cannot track long-term effects, so data mining is required. (Alimova & Tutubalina, 2017). Using social network mining also provides a unique aspect of view. Using social network mining provides a unique aspect of view. There is also a need for studies to observe personal opinions and experiences with chemotherapy (L. Zhang et al., 2018).

2.2.2.4 Difficulties of Interpreting Social Media

One of the major challenges in data mining within HSN is the presence of informally written biomedical information. HSNs often contain a large amount of user-generated content, which includes mistakes, misspellings, and subjective opinions (Miftahutdinov et al., 2017). Online Health Communities (OHC) topics are usually heterogeneous, domain-dependent, and differ from professional terms used in other general-purpose social networks (S. Zhang et al., 2017). The reliability of earlier approaches in data mining is questionable as they tend to overlook essential information. Most studies have focused on the general domain, neglecting the specific needs of the health domain (Na et al., 2012; Sharif et al., 2014). Moreover, previous studies have often disregarded important knowledge, such as semantic information (Sun et al., 2015). Age and gender are critical factors in medical applications, but they are frequently assumed to be available or ignored in previous studies (E. Tutubalina & Nikolenko, 2018). Additionally, understanding user relationships, behaviour, and interactions is crucial for evaluating user survival analysis in HSNs (X. Wang et al., 2014, 2017; Zhao et al., 2014).

2.2.3 Methods

2.2.3.1 Conventional Machine Learning Approaches

The features used in the analysed approaches can be categorised into entity-level and content-level features. Entity-level features include word embedding, cluster-based representation and semantics. Content-level features consist of BOW, parts of speech, sentiment features, pointwise mutual information, medical terms, and emotions (Alimova & Tutubalina, 2017). Among these, the most employed features are semantics (Alimova & Tutubalina, 2017; Cheng et al., 2017; Sharif et al., 2014), BOW (Achrekar et al., 2013; Alimova & Tutubalina, 2017), word embedding (Alimova & Tutubalina, 2017; Sun et al., 2015), sentiment (Alimova & Tutubalina, 2017; Sharif et al., 2014), and emotions. Semantic features, also known as Lexicon features, represent logical entities which are groups of similar keywords (Sharif et al., 2014). BOW is a vector-based method that provides a

simple approach (Achrekar et al., 2013). Sentiment features indicate the polarity of a sentence, whether it is positive or negative, along with the degree of sentiment (Sharif et al., 2014). Word embedding is a technique used to convert words into vectors based on training with existing data. In the reviewed approaches, word embedding can be obtained from an existing model in a social network context (Alimova & Tutubalina, 2017), or generated by clustering indicator seed words, which form the initial dataset (Sun et al., 2015). Emotional features indicate the polarity of sentiment based on subjective words (Sharif et al., 2014). Additionally, (Sharif et al., 2014) also employed a baseline feature that considers the specific meaning of words.

AdaBoost is a classifier that utilises a regression tree as weak learners (Zhao et al., 2014). The boosting method is employed in AdaBoost to combine multiple weak learners into strong learners, although the specifics of how this merging process occurs are not explained. In one application, AdaBoost is applied to predict a user's ability to affect others (Zhao et al., 2014). However, the methodology or approach used for this prediction is not clarified. Additionally, AdaBoost is utilised to categorise five types of social support by extracting post metadata related to lexical, sentiment, and topic features (X. Wang et al., 2014). The specific details on how AdaBoost is applied for this categorisation and how the post metadata is processed and utilised are not provided.

2.2.3.2 Recurrent Neural Network (RNN)

The RNN approaches commonly utilise word embedding in combination with LSTM and CNN. These approaches include a joint LSTM+CNN+CRF method for detecting ADR, a Gated Recurrent Unit (GRU) for detecting medical terms (Elena Tutubalina et al., 2018), and LSTM for indicating account profiles (L. Zhang et al., 2018).

2.2.3.3 Convolutional Neural Network (CNN)

CNN is utilised in data mining of HSN using word embedding. In a study by S. Zhang et al. (2017), a text mining method was proposed that employed a 100-dimensional word vector and CNN for identifying topics of discussion in OHC related to breast cancer. Furthermore, E. Tutubalina and Nikolenko (2018) proposed a CNN approach that utilised the word vector feature for gender and age prediction in the context of different health conditions.

2.2.3.4 Summary of HSN Approaches

Table 2.2 ***Summary of data mining approaches in HSNs.***

Method	**Features**	**Performance (F-measure)**	**Data Size**	**Reference**
SVM	Entity-level (word embedding, cluster-based representation, semantic), content-	80.3%	6320	(Alimova & Tutubalina, 2017)

	level (BOW, parts of speech, sentiment, pointwise mutual, medical terms, emotions)			
	Semantic word count	61.3%	974	(Cheng et al., 2017)
	Collection, unigrams	89.19%	3000	(Sun et al., 2015)
	Baseline, semantic, emotion-related	78.2%	114000	(Sharif et al., 2014)
		79.73%	5000	
	BOW (bag-of-words)	81.4%	25000	(Achrekar et al., 2013)
		76.8%	10000	
	Discriminant word	91.0%	200	(Song & Marsh, 2012)
		80%	50	
Average of SVM		78.8%	164544	
RNN	word level feature	81.0%	875	(Elena Tutubalina & Nikolenko, 2017)
	word embedding	81.41% (accuracy)	1250	(Elena Tutubalina et al., 2018)
	word embedding	85.2% (accuracy)	1824	(L. Zhang et al., 2018)
Average of RNN		83.1%	3949	
CNN	Word vector	65%	33332	(S. Zhang et al., 2017)
	Word vector	50%	33332	(E. Tutubalina & Nikolenko, 2018)
Average of CNN		57.5%	66664	
Logistic regression	Semantic word count	74.8%	394	(Coppersmith et al., 2014)
		62.6%	441	
		76.8%	244	
		57.3%	159	
	Post type	94.3%	1343	(X. Wang et al., 2017)

Average of Logistic regression		82.0%	2581	
AdaBoost	Basic, Lexical, Sentiment, Topic	80%	1333	(X. Wang et al., 2014)
	Lexical	79.2%	298	(Zhao et al., 2014)
Average of AdaBoost		79.9%	1631	
Other Approaches	extract word, part of speech tag, suffix and prefix, content, word type, dictionary lookup, cluster-based feature, word embedding	79.4%	1250	(Miftahutdinov et al., 2017)
	BOW, Negation, type dependencies	0.79	1000	(Na et al., 2012)
Average of Other approaches		79.2%	2250	

From Table 2.2, RNN has an overall better performance and its average F-measure (accuracy) is 83.1% while CNNs have an average F-measure of 57.5%, which is the worst approach. This could be due to RNN being more suitable for sequencing sensitive data. On the other hand, logistic regression and AdaBoost also have around 80% of average performance.

2.3 Generative Large Language Models

Recent studies utilize advanced Artificial Intelligence (AI) technologies, namely Language Models (LLMs) and Retrieval-Augmented Generation (RAG) techniques. LLMs are powerful models that can understand and generate text like human language by learning patterns and structures from large datasets. On the other hand, RAG combines information retrieval and text generation to enhance the performance of AI systems.

2.3.1 Application Areas

The reviewed papers cover a diverse range of application areas within healthcare. These include the analysis of multimodal Electronic Health Records (EHR) to improve clinical predictive capabilities (Zhu et al., 2024), the development of personalized disease prediction LLM system (Jin et al., 2024),

the use of specialized LLMs for liver disease management (Ge et al., 2023), and the implementation of real-time patient-AI healthcare conversations to enabling trust-building and accurate medical support (Mukherjee et al., 2024). Furthermore, the research extends to medical knowledge acquisition and question-answering systems to support healthcare professionals (J. Wang, Yang, Yao, & Yu, 2024), and the exploration of pharmaceutical regulatory compliance in the healthcare industry (J. Kim & Min, 2024).

2.3.2 Problems

The utilization of LLMs in healthcare applications presents several challenges. One challenge is the integration of diverse types of healthcare data, such as time-series data, clinical notes, and knowledge graphs, to improve the prediction of clinical outcomes (Zhu et al., 2024). Another challenge is the integration of vast data volumes and inconsistent symptom characterization standards to provide personalized healthcare (Jin et al., 2024). Furthermore, there is a need for specialized LLMs tailored to specific diseases, like liver disease, to enhance disease management (Ge et al., 2023). LLMs often struggle with the complexity and reliability of medical information in question-answering tasks, indicating the need for advancements in this area. Ensuring safety and reducing hallucinations pose significant challenges for medical LLMs (Miao, Thongprayoon, Suppadungsuk, Garcia Valencia, & Cheungpasitporn, 2024; Mukherjee et al., 2024; J. Wang et al., 2024). Additionally, the utilization of medical knowledge in question-answering tasks presents challenges, including the inefficiency of navigating pharmaceutical guidelines (J. Kim & Min, 2024).

2.3.3 Methods

RAG incorporates a retrieval component to retrieve relevant information and integrates it into the generation process. For example, Zhu et al. (2024) propose the REALM framework, which utilizes LLMs to encode clinical notes and GRU models for time-series EHR data. They also employ RAG to extract task-relevant medical entities for matching in a professional knowledge graph.

In another study, Jin et al. (2024) introduce the Health-LLM system, which combines feature extraction techniques and medical knowledge retrieval using the Llama Index to predict diseases. The Llama Index is employed to retrieve relevant medical knowledge for the prediction process.

Ge et al. (2023) develop a liver disease specific LLM called "LiVersa" by applying RAG on American Association for the Study of Liver Diseases (AASLD) guidelines and documents.

Another approach is presented by Mukherjee et al. (2024), who create a constellation of specialized LLMs. Their primary agent focuses on driving patient-friendly conversations, while several specialist support agents assist with healthcare tasks, such as nurses, social workers, and nutritionists, increase safety and reduce hallucinations.

The Joint Medical LLM and Retrieval Training (JMLR) program trains LLM and retriever jointly, aiming to enhance LLM's ability to leverage medical knowledge for reasoning and answering questions (J. Wang et al., 2024).

J. Kim and Min (2024) also employ a dual-track retrieval approach, utilizing both the user's query and the fine-tuned LLM's answer to expand the search scope and retrieve relevant guideline documents.

Miao et al. (2024) propose a strategy that integrates RAG with a specialized ChatGPT model, specifically tailored to align with the KDIGO 2023 guidelines for chronic kidney disease (CKD).

2.3.4 Result and Summary

The results of these studies demonstrate the potential of AI and LLMs in improving specific aspects of healthcare. For instance, Zhu et al. (2024) show that the REALM framework surpasses baseline models in accurately predicting in-hospital mortality and readmission risks. Jin et al. (2024) demonstrate that the Health-LLM system outperforms existing methods in disease prediction, showcasing its potential for early detection and intervention. LiVersa demonstrates the ability to answer test questions correctly, although there is room for improvement in providing detailed responses (Ge et al., 2023). Polaris demonstrates performance on par with human nurses in a comprehensive clinician evaluation (Mukherjee et al., 2024). Additionally, J. Wang et al. (2024) demonstrate that JMLR models outperform conventional pre-training and fine-tuning approaches in medical question-answering tasks, showcasing their enhanced capabilities. Furthermore, J. Kim and Min (2024) find that the QA-RAG model outperforms other baselines, including conventional RAG methods and other retrieval approaches.

2.4 Trends, Gaps and Opportunities

Table 2.3 ***Summary of Approaches for Healthcare.***

Approach	**Usability**	**Cost Efficiency**	**Efficiency of Medical Services**	**Effectiveness of Medical Services**
HSN	Moderate	High	Low	Low
IoT for Health	Low	Moderate	High	Moderate
Proposed Integrated HSN	High	High	High	High

Through an extensive literature review, this research paper identifies trends, gaps and opportunities in the field of HSNs and Health IoT, highlighting areas for further investigation and improvement.

In terms of trends, there is a noticeable shift towards the utilisation of machine learning and deep learning methods such as CNN, RNN and LSTM. Researchers are employing these techniques for various tasks including classification, topic modelling and information extraction. Additionally, the

use of word embeddings and other NLP techniques have gained traction in analysing the vast amount of textual data generated on social media platforms. These trends signify the growing interest in leveraging advanced computational methods to gain valuable insights from health-related data in HSNs and IoT.

A significant trend in the literature is the increasing application of RAG and LLMs in healthcare. These advanced computational methods are being utilized for a variety of tasks, including the analysis of multimodal EHR, personalized disease prediction, and the development of specialized LLM chat interfaces for conditions like liver disease. The use of RAG and LLMs has gained traction for its potential to refine clinical predictions, enhance feature extraction, and improve the accuracy of AI-generated responses in medical contexts.

However, several gaps still exist in the integration of HSNs and IoT. One significant challenge is the difficulty in interpreting the informal language and grammar used in social media posts, which presents a barrier to harnessing the full potential of these data sources. Furthermore, the reliability and validity of information shared online pose ongoing concerns, making it crucial to develop approaches, such as trustworthiness assessment methods, to ensure the reliability of such data. Ensuring the reliability and validity of information shared online is crucial, necessitating the development of trustworthiness assessment methods to guarantee the accuracy of data utilized by RAG and LLMs.

Despite these challenges, there are several promising opportunities for advancements and improvements in the integration of HSNs and Health IoT. A comprehensive overview of the advantages, disadvantages and proposed approaches is shown in Table 2.3.

HSNs offer a wealth of rich information but concerns regarding privacy and usability persist. On the other hand, IoT solutions provide pervasive and autonomous monitoring capabilities but often require experts to interpret and provide necessary information and care. By integrating data from HSNs and IoT, healthcare providers can obtain a more comprehensive understanding of a patient's health status, enabling them to make informed decisions regarding diagnosis and treatment.

One notable opportunity lies in the development of multimodal approaches that combine data from HSNs and IoT. This integration allows for a more holistic approach to diagnosis and treatment planning, offering healthcare professionals a more complete picture of patients' health conditions. Additionally, leveraging the vast amount of unstructured social data available through HSNs can lead to the generation of structured knowledge, providing valuable insights and support for clinicians. This, in turn, facilitates evidence-based decision-making in healthcare.

Real-time monitoring of health discussions on social media platforms is another opportunity that holds great potential. By closely monitoring these conversations, healthcare professionals can gain early insights into disease trends and outbreaks, enabling more effective prediction and management. Furthermore, personalised recommendation and support systems, driven by social data, could enhance patient care and engagement, tailoring healthcare services to individual needs.

The integration of IoT and RAG-enhanced LLMs presents a significant opportunity to transform healthcare by offering a more holistic approach to patient care. IoT devices are capable of providing objective, real-time physiological and environmental data, which can complement the rich, yet often subjective, information found in social media and clinical notes.

The gap lies in the absence of approaches that effectively combine data from HSNs and IoT to provide comprehensive healthcare solutions. This problem highlights the need to develop methods that bridge the gap between these two data sources.

In conclusion, the integration of HSNs and IoT presents numerous opportunities to revolutionise healthcare. By capitalising on these opportunities, we can significantly improve healthcare outcomes, enhance patient care and address the challenges faced by the healthcare industry.

3 Integrated HSN for Cardiovascular Diseases

3.1 Measuring Similarity between Biosignal with Gaussian Hamming Distance

3.1.1 Introduction

Chronic diseases, such as heart diseases, continue to be a leading cause of global mortality (D Mozaffarian et al., 2016; Dariush Mozaffarian et al., 2015). However, individuals residing in rural and remote areas face significant challenges in accessing healthcare due to high patient-doctor ratios and limited availability of specialised medical equipment (Puustjärvi & Puustjärvi, 2011). As a result, many individuals in these areas often struggle to receive proper diagnosis and treatment for their conditions.

The advent of online platforms has provided an opportunity to address these healthcare access and information gaps. A 2011 survey revealed that 80% of internet users seek healthcare information online (Fox, 2011). HSNs have emerged as a potential solution to improve healthcare access and facilitate remote diagnosis of cardiovascular diseases. HSNs leverage the growing trend of utilising online platforms for healthcare information and support, connecting individuals with similar health conditions, and providing a space for sharing information and support. Through HSNs, individuals in rural and remote areas can benefit from improved access to healthcare resources and specialised knowledge, ultimately enhancing their overall care and wellbeing.

HSNs face challenges in accurately matching patients with similar conditions due to the informal nature of medical descriptions and the presence of spelling errors (Elhadad et al., 2014). Many users encounter difficulties in independently finding suitable online support groups due to their limited expertise in describing and locating relevant communities.

This research aims to overcome these challenges by enabling direct biosignal comparison through Health IoT, facilitating the identification of similar patient groups in HSNs without requiring users to search on their own. By integrating heart sound analysis into HSNs, the research aims to provide heart disease patients with an accessible and reliable method of obtaining objective diagnoses. This integration would improve the overall accessibility and accuracy of healthcare information and support for heart disease patients.

In this section, we propose integrating heart sound analysis into existing HSNs to enable remote diagnosis of heart diseases. Heart sound analysis was integrated into the existing HSNs to enable the diagnosis of heart diseases. A dataset of 76 heart sounds was collected, comprising 18 sounds from healthy hearts and 58 sounds from diseased hearts. Gaussian time series features were extracted from the segmented sounds for subsequent processing. A prediction model was built using this data and a classifier was trained. The Gaussian MFCC achieved the highest accuracy of 84.2% in our experiments.

The contribution is the introduction of Gaussian Hamming Distance (GHD) to identify similar heart condition groups within HSNs. By comparing heart sounds among users within the HSNs, this approach enables the automatic matching of patients with similar conditions based on their heart sounds. This integration of heart sound analysis into HSNs improves the accuracy and convenience of remote diagnosis for cardiovascular disease, eliminating the need for patients to physically visit medical facilities for specialised testing.

In summary, this section addresses the research problem of integrating heart sound analysis capabilities into existing online support groups and health social networks. By exploring the feasibility and potential benefits of incorporating diagnostic features based on heart sounds, the objective is to provide heart disease patients with an accessible and convenient method for obtaining preliminary assessments and identifying potential abnormalities, without the need for in-person medical consultations. The contribution lies in the introduction of the Gaussian Hamming Distance to identify similar heart condition groups within HSNs, thereby revolutionising heart disease management within online communities and facilitating the development of remote healthcare solutions.

3.1.2 Method

3.1.2.1 Gaussian Hamming Distance

GHD (Insu Song, 2015) was applied to the time series data. The input time series was tested with variations such as Time Series (TS), Gaussian Raw Data Time Series (GTS), Gaussian Time Series with Absolute Value (GATS), Gaussian Time Series with Absolute Value and Pooling (GPATS), and Gaussian MFCC Time Series (GMTS).

For GTS, Gaussian was applied directly to the time series of sound data as (3.1).

$$GTS = Gaussian(TS), \tag{3.1}$$

where TS is the input series.

On the other hand, GATS took the absolute value of the raw data instead, as sound data often contains both positive and negative values. This is because sound waves typically exhibit wave-like behaviour, resulting in positive and negative values being near each other. As a result, applying Gaussian directly to sound data usually leads to the loss of sound intensity information, since Gaussian takes the average value of the sound data in its value which includes both positive values and negative values. Therefore, the absolute values of the raw data were utilised as (3.2).

$$GATS = Gaussian(|TS|), \tag{3.2}$$

where *TS* is the input series.

There was another problem encountered during the experiment. The time taken for applying Gaussian filtering and calculating the Hamming distance was found to be significant, particularly when using raw data with a sampling rate of 3000 and a window size of 2 seconds. To mitigate this issue, GPATS was devised to reduce the size of the array. In this method, pooling was applied to the absolute time series data, selecting the maximum value within a specific window length as (3.3).

$$GPATS = Gaussian(pooling(|TS|)), \tag{3.3}$$

where *TS* is the input series.

In the previous approaches, one major drawback was the loss of frequency information. To overcome this disadvantage, the present study utilised the GMTS approach, which applies Gaussian to the MFCC series. In GMTS, Gaussian is applied to the time series of MFCC, incorporating different window sizes, whether overlapping or non-overlapping, without applying Gaussian to MFCC channels as (3.4).

$$\begin{aligned} GMTS = [&Gaussian(ceps{=}0(\ TS\ )),\\ &Gaussian(ceps{=}1(\ TS\ )),\\ &\ldots,\\ &Gaussian(ceps{=}i(\ TS\ ))], \end{aligned} \tag{3.4}$$

where *TS* is the input series, *i* is the number of MFCC channels.

GTS, GATS, and GMTS were evaluated using the K-nearest neighbours (KNN) algorithm and compared with the conventional KNN distance measure approach. The *k* value for the KNN algorithm was determined based on the specificity of the trained model using Leave-One-Out-Validation.

To assess the robustness of the proposed method against environmental noise and inaccurate extraction, a stress test was conducted using Leave-One-Out-Validation. In this test, the testing signal data was shifted along the time axis, either to the left or right. Shifting left involved discarding some data from the beginning of the original sound signal while shifting right involved adding blank padding before the beginning of the original signal. Any excessive signal resulting from right shifting was discarded to maintain the original signal length. The purpose of this stress test was to evaluate the performance of the method under different conditions and assess its resilience to environmental noise and inaccurate segmentation.

Several parameters were considered in the approaches. For all the approaches, the Sigma value of the Gaussian must be set. In the case of MFCC, additional parameters include determining whether there is overlapping of windows, specifying the size of windows, and selecting the number of

channels. When evaluating these parameters, the worst specificity of the stress test was considered, as specificity is a critical factor for the reliability of diagnosis, particularly when dealing with a limited number of negative (healthy) samples.

3.1.2.2 Heart Sound Data Collection

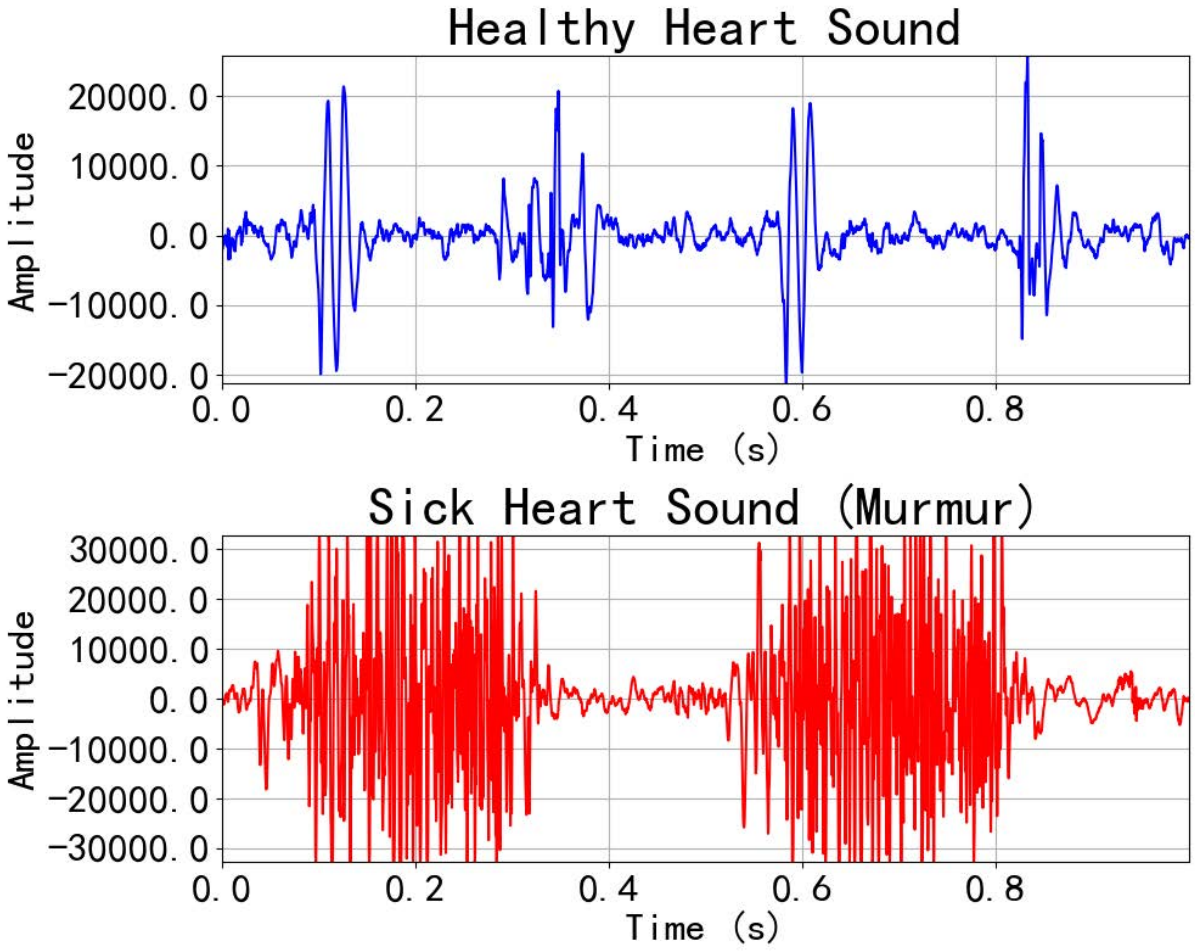

Figure 3.1 ***Example of heart sound in the dataset.***

For the current study, heart sound recordings were retrieved from the internet. A total of 76 heart sound recordings were retrieved from the internet for the current study. Among these recordings, there were 18 instances of healthy heart sounds and 58 instances of heart sounds with various diseases or disorders. Figure 3.1 illustrates examples of collected heart sounds. The upper plot represents a healthy heart sound, characterized by clear S1 and S2 sounds. In contrast, the lower plot depicts a heart with a murmur, where noise is present between S1 and S2. This noise indicates turbulent blood flow through the heart valves or abnormal connections between the heart chambers. The heart disease examples included splitting heart sound, atrial septal defect, click, murmur, regurgitation, Hypertrophic Obstructive Cardiomyopathy, Tricuspid Valve Insufficiency and Aortic Valve disorder. The heart sounds were labelled based on their classification, with healthy heart sounds labelled as Class 0 and abnormal heart sounds labelled as Class 1. To account for variations caused by recording

devices, the audio was normalised to minimise the effects of factors such as recorder sensitivity. Subsequently, heart sounds were selected from the normalised audio recordings. The selected heart sounds were required to exhibit a complete cycle, including both S1 and S2 components.

3.1.2.3 Experiment Setting

We first apply a low-pass filter to eliminate high-frequency noise. Next, we perform segmentation to isolate one complete cardiac cycle. Subsequently, we extract Gaussian time series features from the heart sounds for further processing. Using this data, we construct a KNN prediction model classifier to distinguish between normal and abnormal heart sounds. The number of neighbours (the value of *k*) is one. Our best results were achieved using Gaussian MFCC, which yielded an accuracy of 84.2%.

Two methods have been employed for implementing the KNN algorithm: Hamming distance and Euclidean distance. Hamming distance is used to compare the test data with the training data and is calculated as shown in (3.5). On the other hand, ordinary KNN estimates the distance between vectors using the Euclidean distance, which is calculated as shown in equation (3.6).

$$HD = \sum(x_i - y_i), \tag{3.5}$$

$$ED = \sqrt{\sum(x_i - y_i)^2}, \tag{3.6}$$

where x_i and y_i are each element of two different series.

Another indication of the superiority of GHD was demonstrated by Insu Song (2015). In their study, GHD was able to effectively capture facial images even when working with low-quality images. This highlights the potential of GHD in handling similar challenges in audio signal processing. Like facial images, audio signals can also suffer from low quality. These signals can be captured either manually or automatically. However, both approaches may introduce errors in sound selection, such as including excess blank regions before or after sounds or cutting off parts of the sounds. This shifting of sounds can significantly alter the Hamming distance, resulting in decreased performance. Furthermore, the presence of environmental noise can negatively impact prediction efficiency. The error captured by the Hamming distance can be represented by equation (3.7).

$$HD = E_d + E_r + E_n, \tag{3.7}$$

where difference error E_d is the desired distance for measuring the difference between two sounds while shifting error E_r and noise error E_n should be minimised. To minimise shifting error and noise error, a Gaussian filter is applied to the time series.

Equation (3.8) shows the calculation of Gaussian Hamming distance (GHD) between training data and test data.

$$GHD = |Gaussian(S_{test}) - Gaussian(S_{data})|, \quad (3.8)$$

where S is the input series.

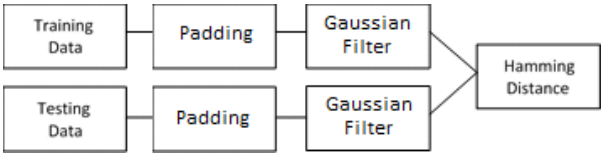

Figure 3.2 ***Process of GHD.***

In the classification procedure for heart sounds using GHD, the heart sound segment S is carefully selected and standardised to a fixed length of 1.5 seconds. To ensure consistency, zero padding is applied to the segment if necessary. The padded signals are then subjected to a Gaussian processing step, where various values of Sigma are utilised. These processed features are subsequently employed in conjunction with the KNN algorithm. Figure 3.2 shows the proposed method for measuring GHD between testing data and testing data.

3.1.2.4 The Evaluation Matrices

The evaluation matrices of our experiments are Sensitivity, Specificity, and Accuracy. Sensitivity, specificity, and accuracy are calculated as:

$$Sensitivity = \frac{TP}{P}, \quad (3.9)$$

$$Specificity = \frac{TN}{N}, \quad (3.10)$$

$$Accuracy = \frac{TP+TN}{P+N}, \quad (3.11)$$

where TP is the number of true positives; P is the total number of positive ground truth cases; TN is the number of true negatives; N is the total number of negative ground truth cases.

Sensitivity and specificity are evaluation metrics that can help evaluate the performance of a model when dealing with imbalanced medical data. They provide valuable information about the model's ability to correctly identify positive and negative cases, regardless of the dataset's imbalanced nature. Sensitivity measures the model's ability to correctly identify positive cases, while specificity measures its ability to correctly identify negative cases. By focusing on these metrics, one can gain insights into the model's performance, independent of the data distribution.

3.1.3 Results

In this section, the results of GTS, GPATS and GMTS are analysed and compared. Their worst possible performance over the stress test is analysed to show the overall performance.

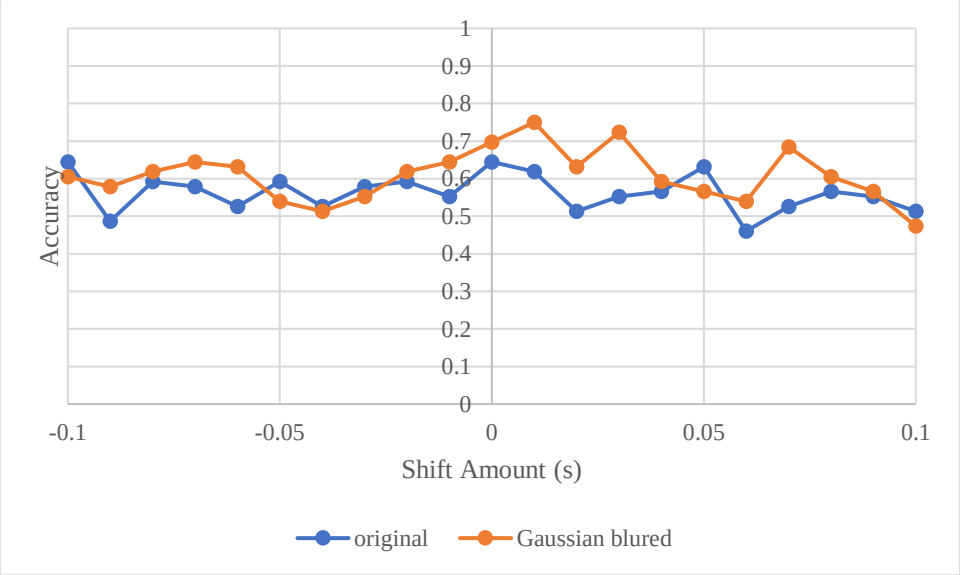

Figure 3.3 ***Accuracy of GTS.***

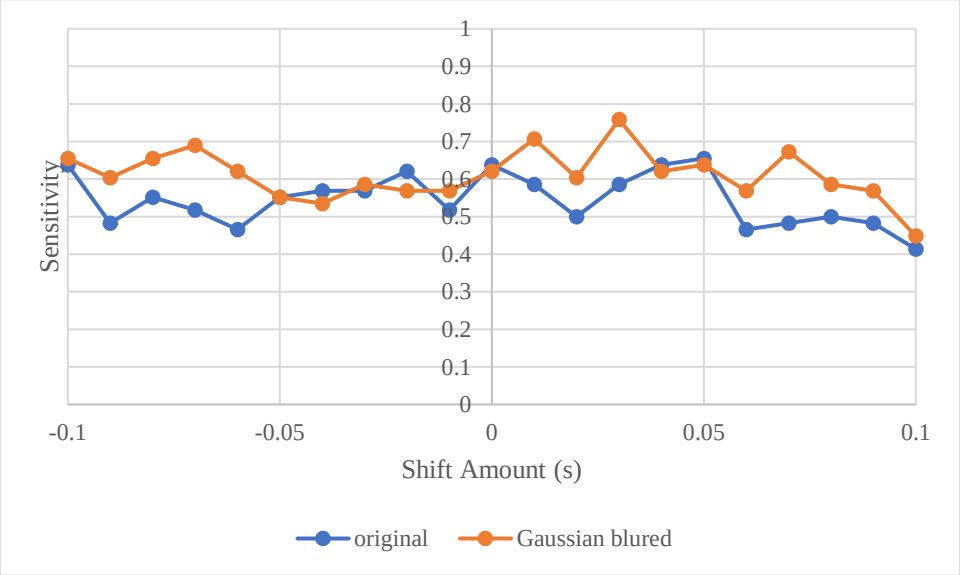

Figure 3.4 ***Sensitivity of GTS.***

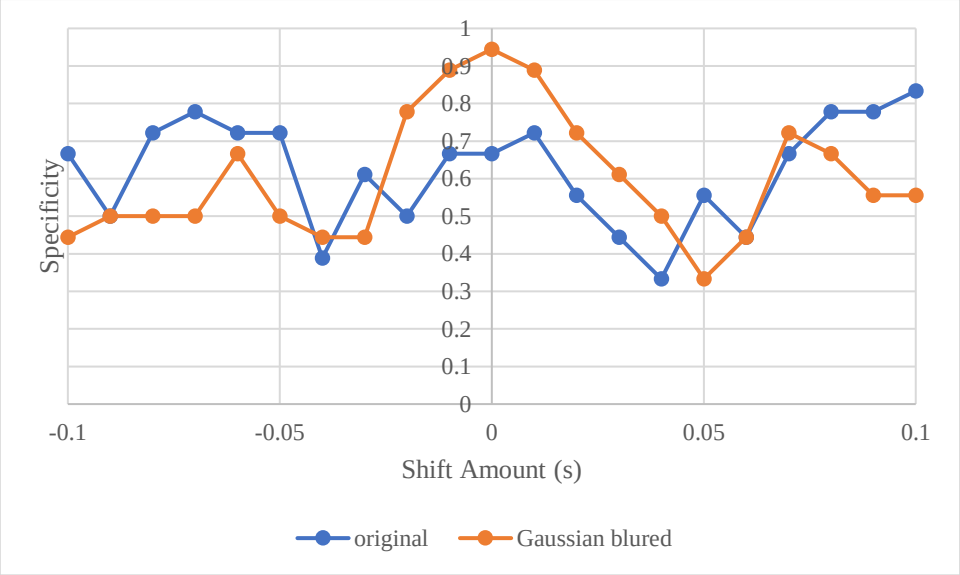

Figure 3.5 ***Specificity of GTS.***

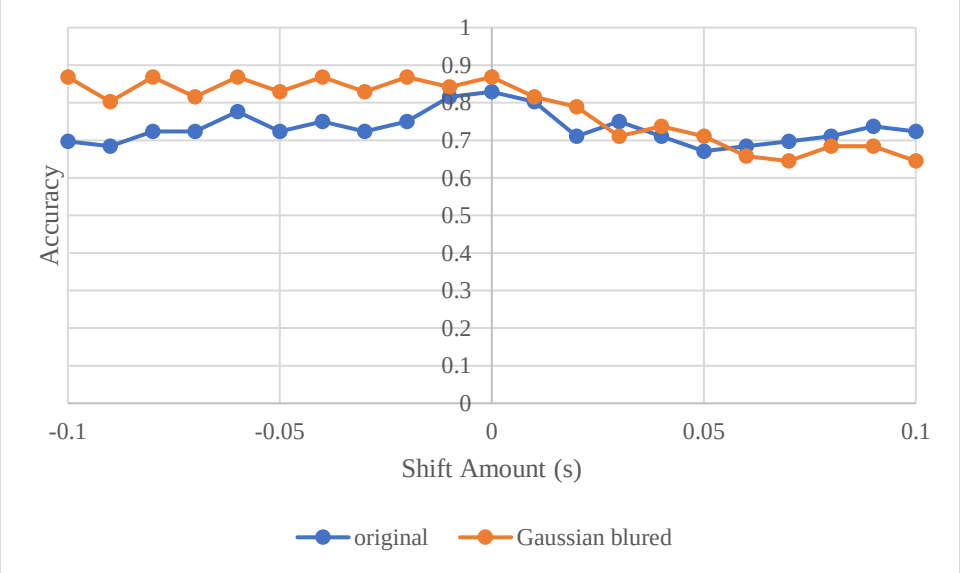

Figure 3.6 ***Accuracy of GPATS.***

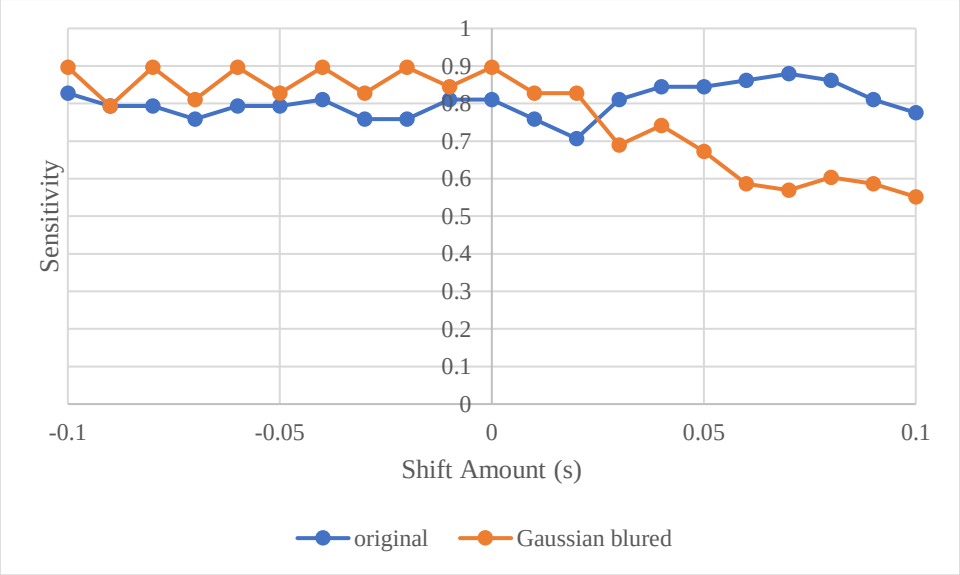

Figure 3.7 ***Sensitivity of GPATS.***

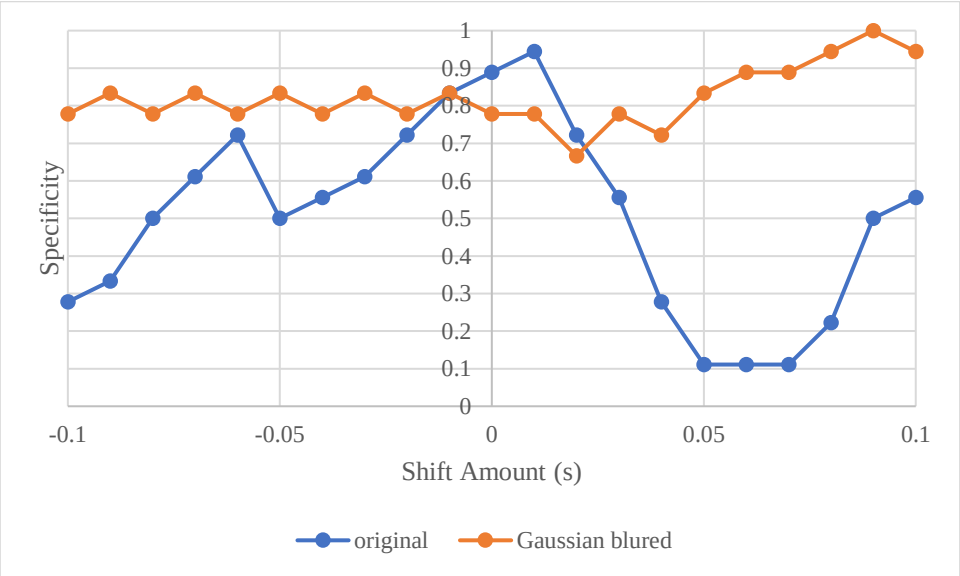

Figure 3.8 ***Specificity of GPATS.***

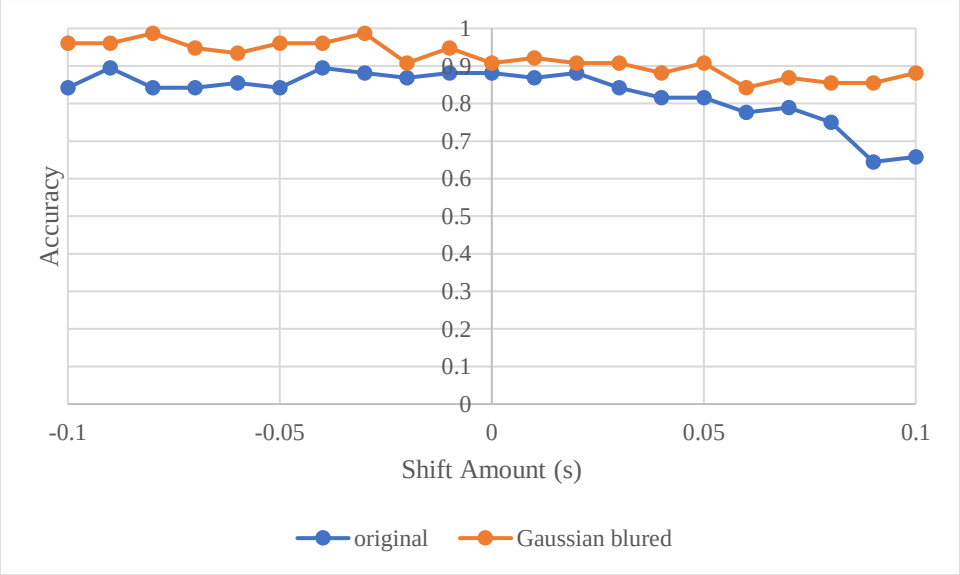

Figure 3.9 ***Accuracy of GMTS.***

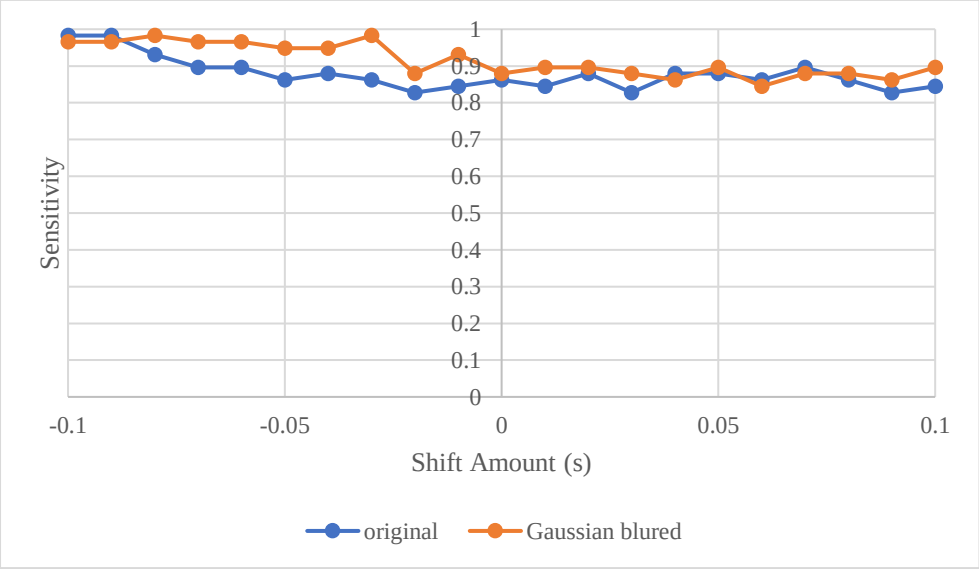

Figure 3.10 ***Sensitivity of GMTS.***

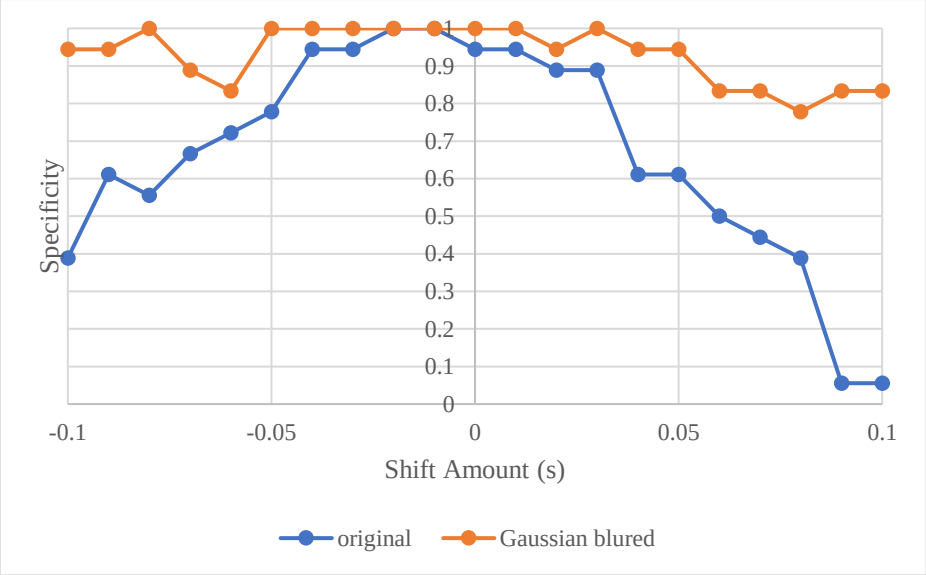

Figure 3.11 ***Specificity of GMTS.***

Table 3.1 ***Minimum accuracy, specificity and sensitivity of all proposed approaches.***

	Accuracy	**Specificity**	**Sensitivity**
GTS	0.4744	0.333	0.448
GPATS	0.632	0.167	0.776
GMTS	0.842	0.778	0.845

Table 3.2 ***Confusion matrix of GMTS.***

	Predicted Positive	**Predicted Negative**
Actual Positive	51	7
Actual Negative	0	18

Table 3.1 presents the minimum accuracy, specificity, and sensitivity of all proposed approaches in the shifting test and Table 3.2 present the confusion matrix of GMTS. The results indicate that GMTS outperforms the other tested methods in terms of accuracy. Additionally, it is observed that applying Gaussian to the time series of most approaches leads to an improvement, except for the sensitivity of GPATS.

Figures 3.3-3.5 display the accuracy, sensitivity, and specificity of GTS. The x-axis represents the amount of shifting in seconds, while the y-axis represents the corresponding performance measurement. It can be seen from the plot that GTS demonstrates significantly higher overall accuracy and sensitivity than the original TS.

However, GTS exhibits reduced specificity when the left shifting amount reaches 0.03s or the right shifting amount reaches 0.05s. This may be caused by the loss of intensity information by applying Gaussian directly to a wave signal. Additionally, there is only a minor improvement in sensitivity,

increasing from 41.4% to 44.8%. Further context is needed to fully understand these percentages and their significance.

Figures 3.6-3.8 show the accuracy, sensitivity, and specificity of GPATS. GPATS proved to be more effective in preserving the intensity information of the sound data and exhibited significantly more stable performance. Figures 3.5-3.7 show the accuracy, sensitivity, and specificity of GPATS. In terms of specificity and sensitivity, GPATS with Gaussian demonstrated more stability and higher performance, although there were slight decreases in some shifting amounts. The minimum specificity achieved by pooling the absolute value of the time series with Gaussian was 66.7%, whereas the minimum specificity in the approach without Gaussian was only 11.1%. Although the sensitivity decreased from 70.7% to 55.2% after applying Gaussian, this can still be considered an improvement since a sensitivity of 55.2% is more acceptable than a specificity of 11.1% for time series without Gaussian. However, despite these improvements, the results were still not satisfactory in terms of sensitivity and overall accuracy.

Figures 3.9-3.11 show the accuracy, sensitivity and specificity of the GMTS algorithm. The GMTS algorithm demonstrated superior performance in terms of overall accuracy, specificity and sensitivity compared with other methods. Specifically, when using the Gaussian MFCC time series, the specificity was more stable, with even minor shifts amounting to 5%. The minimum specificity and sensitivity of the Gaussian MFCC approach were 77.8% and 84.5%, respectively, while the minimum specificity and sensitivity without Gaussian were only 5.6% and 82.8%, respectively. Table 3.2 shows the confusion matrix of GMTS. When comparing GMTS with the normal MFCC approach, which only considers the entire 2-second-length sound wave as its window, significant stability improvements were observed, particularly in terms of specificity, which is crucial for accurate diagnosis.

3.1.4 Conclusion

This section introduces the use of Gaussian Hamming Distance (GHD) for heart sound comparison, addressing research question $RQ_{1.1}$ and proving hypothesis $H_{1.1}$. The challenge faced by users in finding similar patient groups within the HSN using biosignals collected through the Health IoT is highlighted. Users often struggle to locate relevant communities for specific diseases in HSNs and lack expertise in effectively describing their condition, making it challenging to find appropriate communities within the network. This study aims to enable the direct comparison of biosignals through Health IoT, facilitating the identification of similar patient groups in HSNs and overcoming the challenges of finding relevant communities for specific diseases without requiring users to search on their own.

To achieve this, we propose a method that diagnoses heart diseases by utilising the GHD, a similarity comparison method. By comparing the similarity between the heart sounds of individuals and other patients from online support groups, this approach enables accurate searching of online

support groups for certain diseases. In this study, we apply the GHD to data series and time series of features, demonstrating its potential to improve the robustness of the classifier with a shifting range of 5%. This means that applying Gaussian to biosignals could provide the classifier with more tolerance to errors caused by sound selection or differences between different patients. Further tuning of both the Gaussian and the Hamming distance could enhance their tolerance to errors in future research. Additionally, the development of deep learning algorithms is expected to provide more reliable results, and the proposed method can be applied to deep learning in the next stage of research.

The contribution of this study is the introduction of the Gaussian Hamming Distance (GHD) to identify similar groups of individuals with heart conditions within HSNs. By comparing heart sounds among users in HSNs, this approach enables the identification of individuals with similar heart sounds.

Furthermore, the reliable performance of the GHD in handling shifting time series in heart sounds has been demonstrated in previous studies and this study. The proposed method could be utilised in building a system for automatic and early remote diagnosis. With the prevalence of low-cost mobile phones and the extensive development of IoT, a wearable, online and pervasive healthcare system could be developed for various disorders, such as heart disease, gastrointestinal disorders, respiratory diseases and more. These systems have the potential to be diagnosis based on auscultation, providing accessible and efficient healthcare solutions.

3.2 Measuring Similarity Between Biosignal with Predicted Word Embedding Vector

3.2.1 Introduction

Common chronic diseases, such as heart diseases, are the leading causes of death in the world (D Mozaffarian et al., 2016; Dariush Mozaffarian et al., 2015). Continuous care and monitoring are required to protect against these deaths. Health IoT has emerged as a promising approach for pervasive, low-cost, and objective health monitoring (Yildirim, 2018).

ECG is a common biosignal used in Health IoT for diagnosing heart function. ECG provides information on cardiac functions, such as heart rate, and is widely used to monitor cardiovascular disease in IoT applications (Azariadi et al., 2016; Y. Lei, Chungui, & Sen, 2011; Rasid et al., 2014; Verma & Sood, 2018; Yang et al., 2014). The current approaches include real-time ECG beat classification (Tejeda & Falcon, 2019), and heart disease classification (Baloglu, Talo, Yildirim, Tan, & Acharya, 2019; Limam et al., 2017; F. Wang et al., 2007; Xiong et al., 2018; Yildirim, 2018). However, the current Health IoT faces challenges in providing automated search capabilities for relevant health information due to the limited interpretability of biosignal pattern analysis. The currently used automated diagnosis methods, which primarily rely on conventional classification approaches, have inherent limitations that restrict their ability to encompass the wide range of possible

diseases and abnormalities. As a result, manual interpretation of biosignals by trained medical professionals is still required at the end (Manogaran et al., 2018).

To address this gap, we propose the Keyword-based Integrated HSN of Things (KIHoT). The objective of KIHoT is to extend the scope of healthcare data analysis beyond predefined diseases and abnormalities, specifically in the context of cardiovascular diseases and ECG signals. The proposed framework aims to automate the processes of data collection, keyword extraction and information retrieval, thereby enhancing accessibility to healthcare information without the need for expert knowledge.

KIHoT utilises CNN to predict word embeddings of medical keywords from ECG signals. In the previous study (Y. Huang et al., 2017), researchers manually chose the keywords. These medical keywords are selected from HSN posts that are labelled as heart diseases associated with the corresponding ECG signals. The predicted word embeddings enable the framework to search HSNs for relevant medical knowledge. By automating the processes of data collection, keyword extraction and information retrieval, KIHoT aims to enhance accessibility to healthcare information without requiring expert knowledge. The framework enables a direct comparison of ECG signals with textual data within HSNs, allowing for the exploration of healthcare data possibilities beyond predefined diseases and abnormalities.

The proposed method was evaluated using 11,936 ECG signals from patients with heart disease, achieving an average accuracy of 98% for disease prediction. The sensitivity of keyword extraction was found to be more than 90% for all instances and more than 95% for 80% of instances. These results demonstrate the effectiveness of KIHoT in extracting relevant information from HSN portals.

The major contribution of this research is the development of KIHoT which enables direct comparison of ECG signals with textual data for comprehensive access to health information. By integrating IoT and HSNs, KIHoT extends the scope of healthcare data analysis beyond predefined diseases and abnormalities. The framework automates the processes of data collection, keyword extraction and information retrieval, enhancing accessibility to healthcare information without requiring expert knowledge.

3.2.2 Method

3.2.2.1 Overall Process

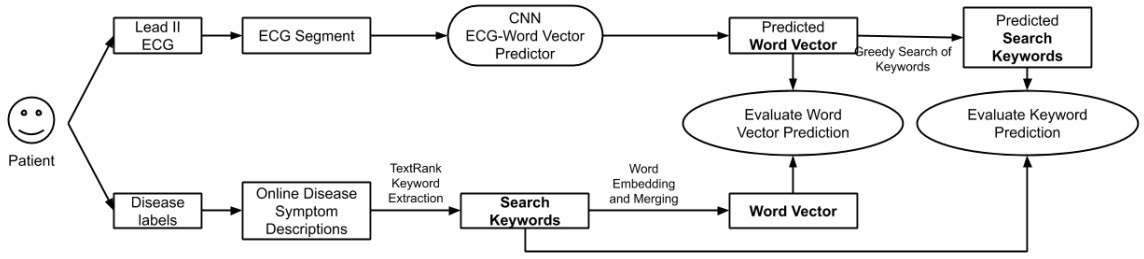

Figure 3.12 ***The overall process of the proposed approach.***

As illustrated in Figure 3.12, the process begins by collecting ECG signals and disease-related expressions from patients. The ECG signals are then segmented and normalised as a preprocessing step for both model training and disease prediction. Symptoms within the disease expressions are selected as keywords, and the labels in the database are transformed into sets of keywords.

Next, the keywords are converted into word embedding vectors, and the sum of these word embedding vectors (SOWE) is calculated. The normalised ECG signals and SOWE vectors are used to train the CNN predictor. The CNN encoder is trained with ECG as input and SOWE as output, enabling it to predict the most likely SOWE given an ECG signal. Subsequently, the predicted keywords are utilised to search for relevant information in HSNs, providing descriptions of the health conditions. The predicted SOWEs are then converted back into keywords to be compared with the original keyword sets. Furthermore, the cosine similarities between the original SOWE vectors are utilised to predict diseases.

3.2.2.2 ECG Data Collection

Since there was no ECG data associated with symptom expressions in HSNs, this study utilised fabricated data to test the proposed approach. A new evaluation model was adopted for this purpose. The data used in this study were collected from the PTBDB, a Physionet dataset. The PTBDB contains 70,207 heartbeats from 268 subjects, along with diagnostic results. The diagnostic labels encompass various conditions such as cardiomyopathy, bundle branch block, hypertrophy, myocarditis, myocardial infarction, valvular heart disease, dysrhythmia and healthy controls. Each recording in the PTBDB comprises 15 signals, including 12 conventional leads and 3 Frank lead ECG. The sampling rate of the signals is 1000.

In the HSNs, patients are not expected to know about their specific diseases. Instead, they are solely expected to be aware of their symptoms and write related posts. Since no dataset currently exists that links condition-related posts with actual diagnosis results, a dataset was created based on the PTBDB. For each disease, a description of its symptoms was selected from the internet and used as the posts. It was assumed that patients with the same disease would exhibit the same symptoms.

Additionally, it was assumed that patients' self-reported descriptions of their conditions were accurate. Hence, keywords were extracted from the symptom descriptions found online, as symptom descriptions are more prevalent in HSNs. The descriptions were obtained from the top search results for each disease.

Key phrases were extracted using the TextRank algorithm (Mihalcea & Tarau, 2004). The top five keywords for each condition were chosen through this approach. For healthy subjects, the keyword "health" was selected instead.

Table 3.3 ***Number of beats in the dataset.***

Disease	Original Dataset	Selected Dataset
myocardial infarction	52326	2496 (5%)
healthy control	10551	2110 (20%)
valvular heart disease	499	499 (100%)
dysrhythmia	1290	1290 (100%)
heart failure	136	-
palpitation	45	-
cardiomyopathy	2227	2227 (100%)
stable angina	162	-
hypertrophy	991	991 (100%)
bundle branch block	2323	2323 (100%)
unstable angina	38	-
myocarditis	475	-

Table 3.3 presents detailed statistics regarding the samples and the distribution of study subjects for each class. The Physionet toolkit was utilised to select the QRS peak. ECG beats were chosen from lead II, as it is widely used in other approaches. Under-sampling was employed to balance the dataset, specifically selecting only 5% of myocardial infarction and 20% of healthy ECG beats from each subject. Some classes with an insufficient number of samples were excluded from this study.

3.2.2.3 The Sum of Word Embedding (SOWE)

The main contribution of this study is the generation of keywords based on ECG data. Instead of using a one-hot label, the model predicts keywords in the form of SOWE (Sum of Word Embedding vectors). The SOWE representation captures both the disease and the bag of keywords. Compared to one-hot annotation, word embedding offers two advantages: it can accommodate a large number of candidate keywords and it preserves the semantic meaning of the keyword set, making disease prediction easier based on a given vector.

In practical applications, determining the exact number of keywords to describe a condition can be challenging. However, with word embedding, a fixed number of attributes can represent a vast number of keywords efficiently. Given the large number of keywords in real-world HSN applications, the word embedding approach demonstrates improved efficiency.

To represent the keywords, an existing word embedding dictionary (Miftahutdinov et al., 2017) was utilised. This dictionary, based on CBOW with 200 attributes, was trained using 2.5 million unlabelled comments from online social networks and scientific lectures. Each word is projected into a 200-dimensional vector. The training of this dictionary was performed using the Gensim library.

To remove unnecessary stop words, keywords were selected from the emulated post dataset, which served as the label. Key phrases were extracted using the TextRank algorithm (Mihalcea & Tarau, 2004), and the top five keywords for each condition were chosen. For healthy subjects, the keyword "health" was selected instead. For each keyword, the corresponding word vector was selected from the existing dictionary mentioned above. Since each disease may have multiple keywords, the keyword vectors were summed up.

The SOWE representation of all diseases forms a matrix for disease diagnosis. Labelling ECG with SOWE directly offers the advantage of easily incorporating new keywords if they appear in the word embedding dictionary. This process is also equivalent to a classification model, where the SOWE vectors are compared with other conditions using cosine similarity measurements, as shown in (3.12), (3.13), and (3.14):

$$Output(X)=W_0 \cdot X+b_0, \tag{3.12}$$

$$Classification(X)=W_1 \cdot Output(X), \tag{3.13}$$

$$Classification(X)=W_1 \cdot W_0 \cdot X+W_1 \cdot b_0, \tag{3.14}$$

where W_0, W_1 are learnable weight, b_0 is learnable bias and X is input.

3.2.2.4 Keyword Extraction from SOWE

To extract keywords from word embedding, a greedy search algorithm (White, Togneri, Liu, & Bennamoun, 2016) was employed in this approach, aiming to extract the most important keywords based on the sum of the word embedding vector. The keyword extraction method consists of two steps: the greedy searching step; and the refining step.

In the greedy searching step, the Bag of Words (BOW) is initialised as an empty set. In each iteration, candidate words that minimise the Euclidean distance between the SOWE of the selected keyword set and the given keyword vectors are selected and added to the keyword set. This process continues until no further keyword can minimise the difference between the SOWE of the keyword set and the given word vectors.

The refining step focuses on replacing each chosen keyword with other candidate words. For each keyword, it is first removed from the BOW. Then, each candidate word is added to the BOW to determine whether it can further minimise the difference. If a candidate word is found to reduce the

difference, it is retained. Otherwise, the refining step is reversed. This refining process continues until there is no further improvement in Euclidean similarity or the maximum repetition limit is reached.

3.2.2.5 CNN ECG-Word Vector Predictor

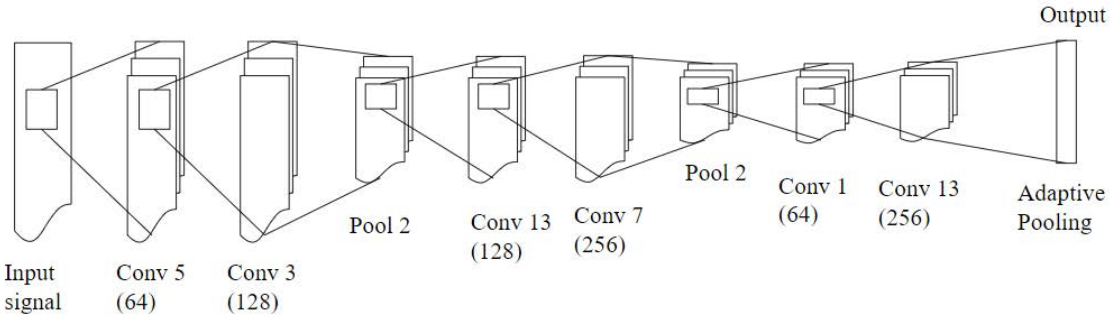

Figure 3.13 ***Structure and Detailed Arguments of The Purposed CNN.***

A CNN is trained in this research to convert an ECG signal segment into a 200-dimensional word embedding vector. Figure 3.13 illustrates the architecture of the CNN utilised in our study. The CNN consists of three convolutional blocks, serving as feature-extracting layers to extract abstract features from the raw ECG signals. These feature extractors are designed to capture fundamental characteristics from preceding layers. The input size of the ECG raw signal is 651, and it is directly fed into the first convolutional block. The final layer of the CNN is a dense layer responsible for regressing the 200 attributes of the word embedding vectors using linear kernels. The loss function employed for regression is a mean square error. As a baseline, another CNN with an identical architecture is trained, but its output layer serves as a conventional multiclass classifier using one-hot labels for keyword prediction. The only difference between the two CNNs lies in their output layers.

Percentage split was the evaluation method in our approaches. We used 70% of the samples for training and 30% of the samples for evaluation. We trained two different CNNs with 50 epochs and batch sizes of 32. The model is trained with SGD (Stochastic Gradient Descent) with learning rate of 0.0001. The first CNN was the proposed approach, which predicted the keyword vector of a condition given ECG. The other CNN has an output layer as a conventional multiclass classifier, serving as a benchmark.

3.2.2.6 Evaluation Matrices

The proposed method was evaluated through three different approaches. The first evaluation focused on measuring the model's ability to predict word embedding vectors. The second evaluation aimed to assess the accuracy of retrieving keywords from the predicted vectors. The third evaluation involved measuring the difference between the predicted vectors and target vectors. To quantify this difference, a classification task was performed using the diseases as classes. For each predicted vector, the most similar vector from the set of conditions was selected. The condition associated with the selected

vector was then compared with the actual condition for evaluation of the prediction. This evaluation task aimed to assess the likelihood of confusion among the conditions.

To further evaluate the advantage of using SOWE over the one-hot label approach, leave-one-class-out testing was conducted. This testing procedure involved completely removing one class from the training set while still attempting to predict that class in the testing set. This evaluation method assessed the ability of the proposed model to predict unknown diseases that were not present in the training data.

3.2.3 Result

3.2.3.1 Keyword Vector Prediction Result

Table 3.4 ***Training MSE, Testing MSE and Keyword Extraction Accuracy of CNN and Linear Regression.***

	CNN	**Linear Regression**
Training MSE loss	0.276	10.154
Testing MSE loss	0.468	12.903
Keyword extraction accuracy	98.604%	57.498%

Table 3.4 presents the training Mean Squared Error (MSE), testing MSE and keyword extraction accuracy for both CNN and linear regression models. The CNN model demonstrates significantly higher performance compared to linear regression.

To assess the usefulness of the proposed approach, an evaluation was conducted by measuring the accuracy of extracting keywords using the predicted vectors. The percentage of correctly extracted keywords was compared to the original keywords. The predicted keywords were obtained using the greedy search algorithm. For instance, an ECG signal associated with a certain condition, such as Myocardial Infarction (MI), would produce a 200-dimensional vector.

Subsequently, the extracted keyword is selected from a candidate keyword dictionary. This dictionary consists of symptoms related to all known diseases and is prepared for the keyword selection process. The significant difference in MSE values leads to varying levels of keyword extraction accuracy. The CNN model achieves 98.6% accuracy in extracting keywords correctly, whereas linear regression only achieves 57.5% accuracy.

The sensitivity is calculated as (3.15):

$$Sensitivity = \frac{number\ of\ correct\ predicted\ keywords}{number\ of\ ground\ truth\ keywords} \tag{3.15}$$

while the precision is calculated as (3.16):

$$Precision = \frac{number\ of\ correct\ predicted\ keywords}{number\ of\ predicted\ keywords} \tag{3.16}$$

The keywords, such as pain, dizziness, rhythms, and vomiting, are selected using a greedy search algorithm from the extracted vector. For example, the ground truth keywords for MI, include pain, dizziness, weakness, heaviness and vomiting. In this example, three out of the five ground truth keywords are correctly predicted, resulting in a sensitivity of 60%. On the other hand, out of the four keywords selected by the algorithm, only three are correct, leading to a precision of 75%.

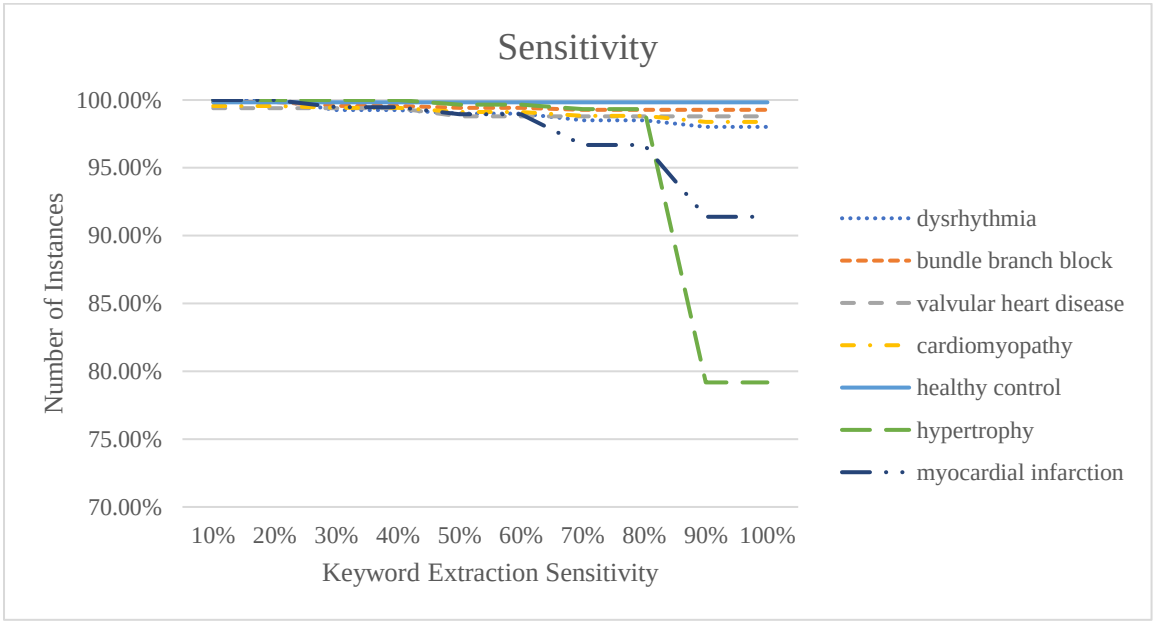

Figure 3.14 ***Sensitivity of keyword extraction of CNN.***

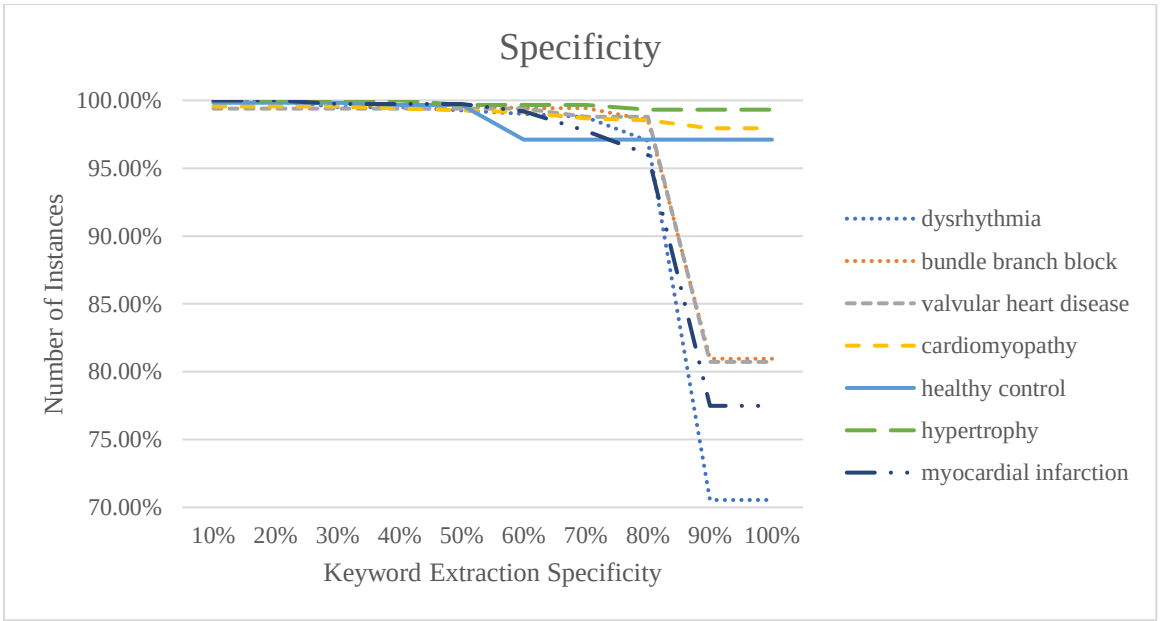

Figure 3.15 ***Specificity of keyword extraction of CNN.***

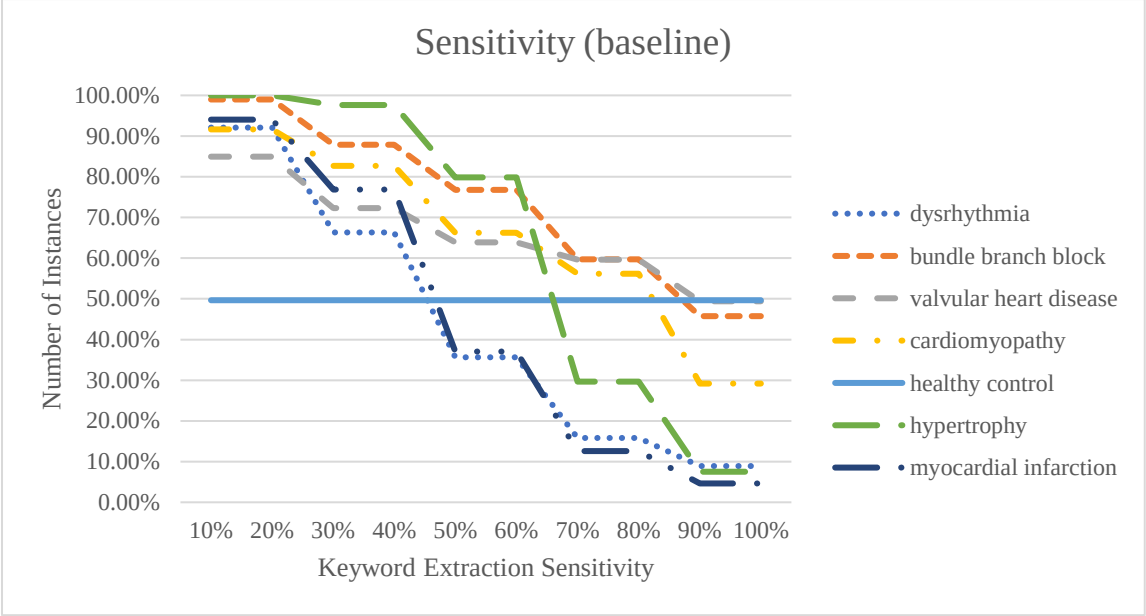

Figure 3.16 ***Sensitivity of keyword extraction of linear regression.***

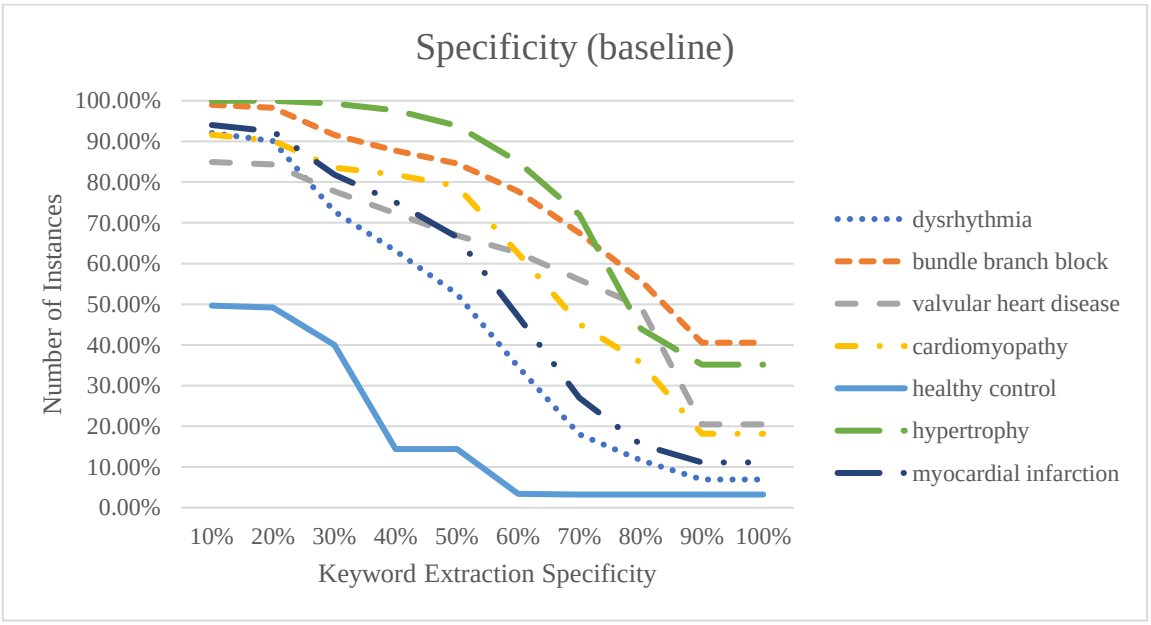

Figure 3.17 ***Specificity of keyword extraction of linear regression.***

Figure 3.14 (scale from 70% to 100%) displays the number of instances with different sensitivity of keywords correctly predicted. The results show that our model successfully extracted all possible correct keywords for more than 70% of instances, and it achieved more than 80% accuracy for more than 95% of instances. Except for the hypertrophy and health control conditions, where the extraction rate was lower, our model correctly identified more than 95% of all keywords for most conditions. Therefore, our model provided sufficient information for most instances. In contrast, the baseline approach, as depicted in Figure 3.16 (scale from 0% to 100%), only provided sufficient information for a limited number of instances.

Figure 3.15 (scale from 70% to 100%) illustrates the population of instances that met various performance standards for extracting valid keywords. These standards represent the sensitivity of keyword extraction. The results show that the sensitivity of keyword extraction was more than 90% for all instances and exceeded 95% for 80% of the instances. In comparison, the baseline method, as shown in Figure 3.17 (scale from 0% to 100%), exhibited much lower sensitivity and only provided useful keywords for a few conditions.

Table 3.5 ***Average performance of keyword extraction.***

	Sensitivity	**Specificity**
dysrhythmia	98.960%	94.345%
bundle branch block	99.481%	96.403%
valvular heart disease	99.036%	96.185%
cardiomyopathy	99.062%	99.005%
healthy control	99.830%	98.441%
hypertrophy	95.631%	99.710%
myocardial infarction	97.298%	95.327%

Table 3.6 ***Average performance of keyword extraction (baseline).***

	Sensitivity	**Specificity**
dysrhythmia	43.762%	46.863%
bundle branch block	73.824%	76.047%
valvular heart disease	66.024%	61.228%
cardiomyopathy	65.191%	62.467%
healthy control	49.660%	19.694%
hypertrophy	62.935%	78.188%
myocardial infarction	45.033%	54.503%

Table 3.5 and Table 3.6 present the average sensitivity and specificity of keyword extraction for both the CNN approach and the linear regression approach. When comparing the performance of these approaches, the CNN approach demonstrated superior results in terms of both sensitivity and specificity for all conditions. To be more specific, the CNN approach consistently outperformed the linear regression approach in accurately identifying keywords. In addition, our deep learning approach provided more stable results and yielded more useful information when compared with conventional machine learning approaches. The CNN approach exhibited a larger proportion of instances with high sensitivity and specificity, indicating its effectiveness in keyword extraction.

3.2.3.2 Disease Prediction Result

Table 3.7 ***Classification Result by CNN Classifier with Softmax.***

	Accuracy	**Sensitivity**	**Specificity**
dysrhythmia	90.347%	96.053%	99.528%
bundle branch block	98.268%	94.452%	98.615%
valvular heart disease	84.337%	97.902%	99.912%
cardiomyopathy	98.974%	89.286%	97.206%
healthy control	90.476%	93.662%	98.797%
hypertrophy	87.372%	96.241%	99.696%
myocardial infarction	89.139%	90.094%	97.381%
meta-average	91.273%	93.956%	98.734%

Table 3.8 ***Searching Performance of word Embedding predicted by CNN.***

	Nearest	2nd	3rd
dysrhythmia	99.505%	99.505%	100.000%
bundle branch block	99.711%	100.000%	100.000%
valvular heart disease	98.795%	99.398%	99.398%
cardiomyopathy	99.413%	99.413%	99.707%
healthy control	99.490%	99.490%	99.830%
hypertrophy	99.317%	99.317%	99.659%
myocardial infarction	99.603%	99.735%	99.735%
meta-average	99.405%	99.551%	99.761%

*Table 3.9****Searching Performance predicted by linear regression.***

	Nearest	2nd	3rd
dysrhythmia	57.426%	76.238%	91.337%
bundle branch block	92.496%	98.990%	99.567%
valvular heart disease	63.253%	68.675%	78.313%
cardiomyopathy	78.299%	81.671%	86.217%
healthy control	13.095%	18.027%	25.510%
hypertrophy	78.157%	93.857%	98.294%
myocardial infarction	62.384%	81.192%	87.550%
meta-average	63.587%	74.093%	80.970%

Table 3.10 ***Searching Performance (Leave-One-Class-Out).***

	Nearest	2nd	3rd
dysrhythmia	0.000%	27.228%	81.931%
bundle branch block	0.000%	13.131%	71.573%
myocardial infarction	0.000%	9.536%	48.079%

The classification performance of word embedding with a linear kernel was compared with a baseline classification method. To perform classification tasks using word embedding, cosine similarities between the predicted vector and the sum of keyword vectors for each disease are calculated. Table 3.7 presents the classification results of the CNN classifier, while Tables 3.8 and 3.9 showcase the classification results of a cosine similarity-based approach for searching. Specifically, Tables 3.8 and 3.9 evaluate the performance of classifiers in identifying the correct condition within the top *N* number of nearest conditions based on the cosine similarity of word vectors. Our findings indicate that the cosine similarity-based classification achieved slightly better performance compared with the CNN classifier. Furthermore, when considering the linear regression approach, most conditions still demonstrated acceptable performance when considering the top two or top three similar conditions. This suggests that the sum of the symptom keyword vector had a positive impact on the classification and diagnosis of diseases.

Table 3.10 demonstrates that using SOWE as the label has the potential to predict diseases that are not present in the training set. Unlike existing disease predictors that output a finite number of labels, our proposed approach predicts a word embedding vector that is associated with symptoms. This means that theoretically, our approach can predict symptoms related to new diseases that were not included in the training data, if the new disease and its related symptoms are provided. To assess the effectiveness of predicting symptoms for unseen diseases, we conducted leave-one-class-out validation. In this validation, each disease was withheld during training to create a model, which was then used to predict the held-out disease. Despite the potentially significant differences in keyword sets among different diseases, SOWE effectively captures the semantics of these keyword sets. Word embedding enhances the performance of the CNN model and yields acceptable results with linear regression. This improvement demonstrates that the embedding dictionary contains valuable information about disease similarities, allowing models to establish connections between signals and embeddings. Our method can be considered a zero-shot learning approach, as it learns semantic embeddings of ECG rather than specific disease classes. Consequently, these semantic embeddings have the potential to predict unknown diseases based on their descriptions.

Table 3.11 ***Comparison with state-of-the-art ECG deep learning approaches.***

Approach	**Task**	**Performance**
CRNN (Limam et al., 2017)	Atrial Fibrillation Detection	90.6% Score

Deep BiLSTM (Yildirim, 2018)	ECG Classification	100.0% Macro F1 score
CNN (Liu et al., 2021)	Arrhythmia classification	98.96%
Self-supervised representation learning with LSTM+MLP (Multiple Layer Perceptron) (Mehari & Strodthoff, 2022)	Disease classification	94.18%
SOWE regression with CNN (Ours)	Disease classification	99.40%

Table 3.11 compares our approach with the recent state-of-the-art ECG deep learning approaches. Our approach shows outstanding performance compared to both supervised classification and self-supervised learning.

3.2.4 Discussion

The results demonstrate that when predicting diseases using the sum of word embeddings instead of one-hot annotation of Bag of Words, the distributed representation is more robust. The sum of the keyword vector offers a simple approach for estimating diseases given keywords and has the potential to predict unknown diseases do not present in the training set. This scalable diagnostic framework leverages the information in HSNs and demonstrates the effectiveness of using the sum of word embeddings to differentiate diseases based on their symptoms.

With an average accuracy of 98% for disease prediction, KIHoT holds promise as a valuable tool for researchers, clinicians and patients. By integrating HSNs and IoT, KIHoT offers an innovative method for collecting data and labels without requiring expert knowledge. The proposed system enhances accessibility to healthcare information while simplifying the data collection and analysis processes. KIHoT's success in predicting relevant keywords from ECG signals positions it as a promising solution for monitoring health conditions and providing timely interventions. Overall, we believe that KIHoT has the potential to significantly improve healthcare services by making reliable health information more easily accessible to all users.

While this study offers a feasible solution for integrating HSNs and IoT to provide cost-efficient healthcare services, some limitations require further investigation. These include the accuracy of the disease-related keyword prediction model, the quality of extracted keywords, and the word embedding dictionary. Additionally, the present study utilised emulated data, which may limit the applicability of the proposed approach in real social network environments. Further research should focus on improving keyword extraction methods to enhance performance and explore the need for a more robust model. Future studies should also aim to expand KIHoT's capabilities to cover a wider range of diseases and assess its scalability with larger datasets from multiple sources.

In conclusion, this study introduces KIHoT, an automated diagnosis framework that integrates HSNs and IoT to address challenges such as rising healthcare costs, limited accessibility to healthcare information, and the difficulty of finding relevant medical knowledge in HSNs. KIHoT utilises ECG

signals to predict disease-related keywords and enhance accessibility to healthcare information. The study demonstrates that KIHoT effectively extracts relevant information from HSN portals, achieving an average accuracy of 98% for disease prediction and providing a substantial number of valid keywords.

3.2.5 Conclusion

This section introduces KIHoT, a deep learning framework that integrates IoT and HSNs, solves research question $RQ_{1.2}$ and proves hypothesis $H_{1.2}$. This study addresses the challenge of converting ECG signals into keywords that allow for direct comparison with textual data in HSN searches. Conventional ECG classification methods are limited to predicting predefined diseases and abnormalities from their training datasets, which restricts their coverage of the wide range of possibilities not explicitly trained for. This study aims to enable the direct comparison of ECG signals with textual data in HSN searches, expanding the possibilities of healthcare data beyond predefined diseases and abnormalities.

To achieve this, we propose a Keyword-based Integrated Health Social Network of Things (KIHoT), a system that predicts word embeddings of medical keywords from ECG signals. KIHoT utilises CNN for this prediction, enabling the search of relevant medical knowledge within HSNs and facilitating a more comprehensive exploration of health information.

The main contribution of this research is the proposal of KIHoT, which enables the direct comparison of ECG signals with textual data for comprehensive access to health information. One significant aspect of KIHoT is its ability to predict word embeddings of medical keywords from ECG signals using CNN. This prediction allows for the search of relevant medical knowledge within HSNs, providing a more thorough exploration of health information.

Furthermore, the effectiveness of KIHoT is demonstrated through its evaluation using a substantial dataset of ECG signals from patients with heart disease. The framework achieves high accuracy in disease prediction and exhibits a high sensitivity in keyword extraction. These results highlight the efficacy of KIHoT in extracting relevant information from HSN portals.

The proposed KIHoT system provides a cost-efficient method for health monitoring by automating data collection, data labelling and model training processes. It requires no expert knowledge from users, significantly expanding accessibility to healthcare information. This approach focuses on helping users search for information about their health conditions on the internet. Based on the highly active user base in HSNs, the integrated HSNs could handle a wider range of diseases compared with normal automated diagnosis approaches, further enhancing its usefulness and impact in the field of healthcare.

3.3 CardioVec: Searching and Indexing ECG in HSN

3.3.1 Introduction

HSNs provide a scalable, sustainable and rich medical knowledge base. IoT provides a non-invasive, easy, low-cost way to collect patient data. However, retrieving the right information from HSNs can be time-consuming and challenging as users are often required to use the right keywords to search and filter relevant information. The large amount of data from HSNs is the largest barrier to users searching for relevant information. HSNs rely on user-entered keywords to find the right information. Finding proper keywords to search the internet for diseases requires in-depth experience in internet searching and medical knowledge. Due to user inexperience and expertise, search results of HSNs often contain errors and informal terms, thus leading to inaccurate analysis (Alimova & Tutubalina, 2017; Miftahutdinov et al., 2017).

There are many applications of IoT focus on disease diagnosis and prediction, such as heart disease early detection (Kumar & Gandhi, 2018; Verma & Sood, 2018) and monitoring (Azariadi et al., 2016; Manogaran et al., 2018). Other approaches include an automatic medicine dispenser (Yang et al., 2014) and community health monitoring (Adame et al., 2018; Jara, Zamora-Izquierdo, & Skarmeta, 2013; Y. Lei et al., 2011; Mdhaffar, Chaari, Larbi, Jmaiel, & Freisleben, 2017; Rahmani et al., 2018). ECG provides information on cardiac functions, such as heart rate, and is widely used to monitor cardiovascular disease in IoT applications (Azariadi et al., 2016; Y. Lei et al., 2011; Rasid et al., 2014; Verma & Sood, 2018; Yang et al., 2014). The current approaches include real-time ECG beat classification (Tejeda & Falcon, 2019), and heart disease classification (Baloglu et al., 2019; Limam et al., 2017; F. Wang et al., 2007; Xiong et al., 2018; Yildirim, 2018).

The current Health IoT faces challenges in providing automated search capabilities for relevant health information due to the absence of interpretable biosignal pattern analysis. This hinders effective searches for health-related information within HSNs, requiring manual interpretation and monitoring by trained medical professionals (Manogaran et al., 2018). Existing ECG diagnostic methods still require experts to interpret and monitor ECG data at the end. This is costly and time-consuming.

This research aims to enable semantic search of text-based content in HSNs using ECG data from Health IoT. The proposed method focuses on converting biosignal data into clinically interpretable descriptions, which can serve as keywords for searching relevant health information in HSNs. Thus, instead of users manually interpreting and entering keywords, the HSN content can be searched using automatically generated keywords from the ECG signals.

To address these challenges, we have developed CardioVec, a comprehensive framework comprising four modules: the ECG wave detector, the ECG physiological measurement extractor, the ECG abnormality detector and the IoT-HSN search interface. The ECG wave detector identifies and locates specific positions of the ECG wave signals, from which physiological measurements are

extracted. These measurements undergo further processing to generate human-readable keywords. Finally, SBERT is applied to search relevant information from the internet using the generated clinical descriptions as keywords.

In our evaluation, CardioVec achieved impressive results, with a top-one precision of 79% for all types of ECG signals, 75% for myocardial infarction ECG signals, and 97% for cardiomyopathy ECG signals. These findings demonstrate the effectiveness and potential of CardioVec in enabling the semantic search of ECG data within HSNs, facilitating efficient access to relevant health information for improved decision-making in healthcare.

The main contributions of this work are as follows. First, this study provides a feasible solution for leveraging HSNs with a novel IoT approach that translates biosignals into human-readable keywords. Secondly, the integration of IoT and HSNs allows doctors and patients to search relevant medical information directly using ECG biosignals. Our keyword generation is evidence-based and is more readable and more informative for doctors and patients. The proposed system aims to provide HSN services from end-to-end that do not require any expert knowledge.

3.3.2 Method

3.3.2.1 Overall Process

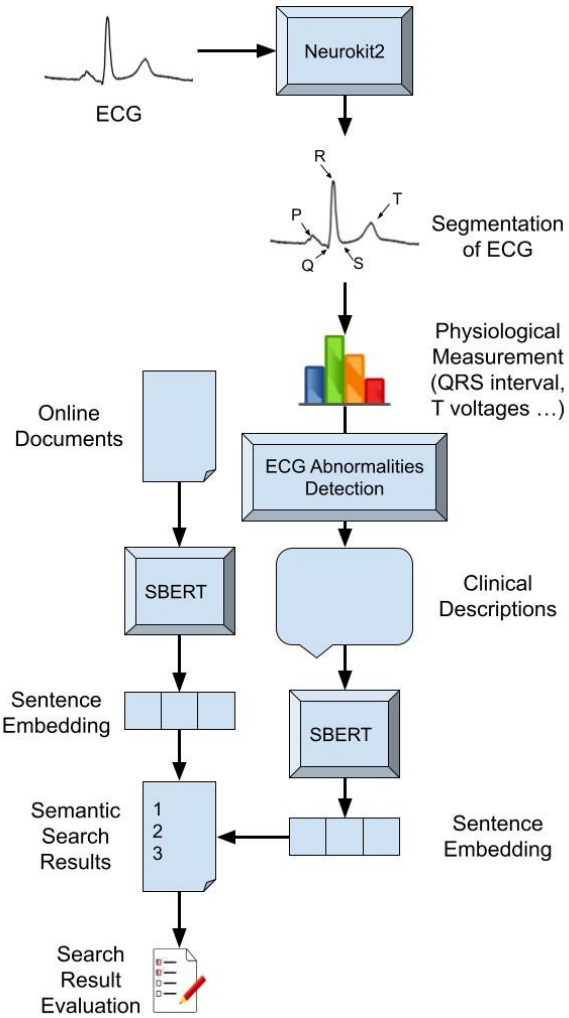

Figure 3.18 ***Overall process of proposed method.***

Figure 3.18 shows the overall process of the proposed approach. CardioVec consists of four modules: ECG wave detector, ECG physiological measurement extractor, ECG abnormality detector and IoT-HSN search interface. The wave positions of ECG signals are detected using the existing toolkit called "Neurokit2" (Makowski et al., 2021), which also annotates PQRST waves on each ECG signal segment. The ECG physiological measurement extractor calculates five types of physiological measurements for each ECG signal, including intervals and voltages. The ECG abnormality detector predicts the presence of six types of ECG abnormalities. Finally, SBERT is applied to search relevant information from the internet using the generated clinical descriptions as keywords.

3.3.2.2 Dataset

The ECG data was collected from PTBDB, a dataset containing eight diseases from 268 subjects with diagnostic results (Bousseljot, Kreiseler, & Schnabel, 1995). Each PTBDB recorded 12 conventional

leads and three Frank lead ECGs. We used ECG pulses from Lead II because Lead II is widely used in other approaches. Some ECG recordings were too noisy to extract the annotations with the existing toolkit, so we discarded them. In this study, we focused on myocardial infarction, cardiomyopathy, heart failure and angina. Table 3.12 shows the disease class, detailed numbers of ECG samples, and number of articles retrieved.

Table 3.12 ***Number of ECG and Articles Collected for each Disease.***

Diagnostic Class	Number of Selected ECG Subjects	Number of Collected Articles
myocardial infarction	146	9
cardiomyopathy	28	10
heart failure	3	10
angina	2	10

We collected 39 articles from the internet. The name of each disease was searched on Google and the results on the first page were collected. We collected those articles on diseases to evaluate whether our clinical descriptions had the potential to categorise each disease even without directly predicting the disease name. For each article, the main contents (usually in "<p>" and "<ul>" tags) were extracted. The contents were then converted into plain text and labelled with the disease names.

3.3.2.3 ECG Segmentation

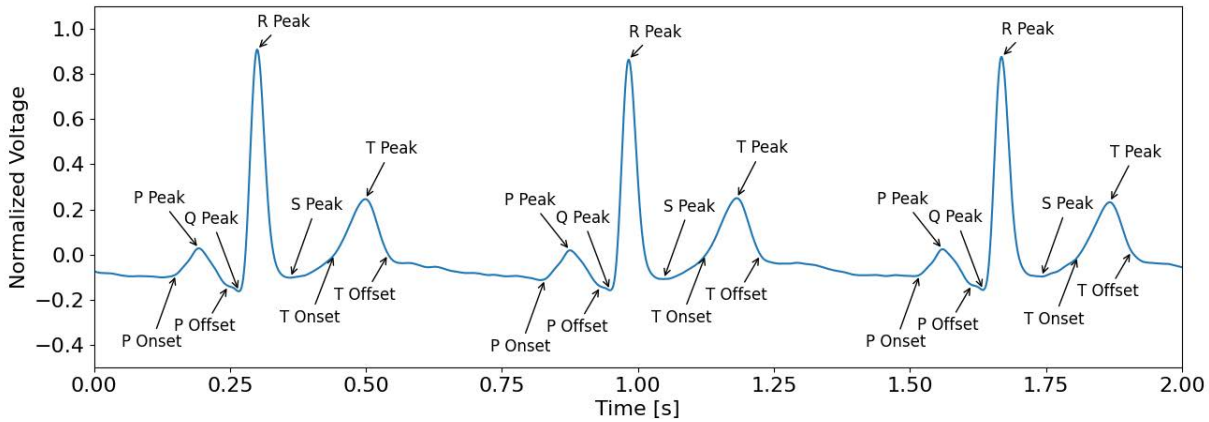

Figure 3.19 ***PQRST waves in a lead II ECG heartbeat.***

Wave positions of ECG signals were detected with "Neurokit2" (Makowski et al., 2021). ECG was cleaned before detection of R peaks (Pan & Tompkins, 1985). To extract the peaks of ECG, the positions of R waves were first extracted. Then the relevant peak positions in the same heartbeat were extracted: P wave, Q wave, S wave and T wave. Finally, the onset and offset of P waves and T waves were also extracted. Figure 3.19 illustrates an example of the segmentation result in a lead II ECG from the PTBDB dataset using "Neurokit2". The PQRST waves' peaks are labelled, and the start and end offsets of the P and T waves are also marked.

ECG physiological measurements were extracted from the interval and voltages of annotations. We extracted five types of physiological measurements, listed as follows.

3.3.2.4 ECG Physiology Measurement

The QT interval is the interval between the peak of the Q wave and the peak of the T wave as in (3.17).

$$I_{QT} = P_T - P_Q, \tag{3.17}$$

where P_T denotes the position of T peaks, and P_Q denotes the position of Q peaks.

QRS interval is the interval between the position of the Q wave and the position of the S wave as in (3.18).

$$I_{QRS} = P_S - P_Q, \tag{3.18}$$

where P_S denotes the position of S peaks, and P_Q denotes the position of Q peaks.

ST voltage measures the voltage of ECG pf ST segment as in (3.19).

$$V_{ST} = ECG(P_{T0}), \tag{3.19}$$

where P_{T0} denotes the onset of T peaks located in the middle of the ST segment.

T voltage measures the voltage of ECG in the T peak as in (3.20).

$$V_T = ECG(P_T), \tag{3.20}$$

where P_T denotes the position of T peaks.

3.3.2.5 ECG Abnormal Patterns

Table 3.13 ***Clinical description definitions.***

ECG Abnormalities	**mean I_{QT}**	**mean I_{QRS}**	**min V_T**	**min V_{ST}**	**max V_{ST}**
ST depression				< -0.1	
ST elevation					> 0.1
Prolonged QRS interval		> 0.12			
Short QRS interval		< 0.1			
Prolonged QT interval	< 0.375				
T wave inversion			< 0		

From the ECG, we extracted six abnormal ECG patterns: normal sinus rhythm, ST depression, ST elevation, sinus tachycardia, sinus bradycardia and atrial fibrillation. Table 3.13 shows how ECG signals can be classified into each type of abnormal heart rhythm.

ST depression and ST elevation suggest angina and myocardial infarction. ECG signals of ST depression have V_{ST} voltage values that are less than -0.1mV. ECG signals of ST elevation have V_{ST} voltage values that are greater than 0.1.

Prolonged QRS interval and short QRS interval suggest cardiomyopathy. The prolonged QRS interval is the average QRS interval over 120ms, and the short QRS interval is the average QRS interval over fewer than 100ms.

The prolonged QT interval is the average QT interval over fewer than 375ms. This value is 2 Sigma away from the average QT interval of healthy subjects.

T wave inversion suggests myocardial infarction. T wave inversion is determined by the voltage of the T wave peak. Normal T wave in lead II is upward and has a positive voltage. A negative voltage suggests an inverted T wave. To detect an ECG recording with at least one inverted T wave, we calculated the minimum voltages of T peaks.

3.3.2.6 Semantic Searching

Embedding is a vector projection space, commonly used for comparing semantic similarity in NLP. For example, phases that have similar meanings such as “breath” and “respiratory” have similar values in embedding. The previous approach (Y. Huang & Song, 2019), used the SOWE to represent the disease by calculating the SOWE of the BOW. The problem with using SOWE is that BOW does not contain word series information, so it does not fully represent the full semantic information of the sentence. Compared with SOWE, SBERT dynamically updates the integration of frame-based word pieces to learn the full semantic information of the sentence.

However, the conventional approach uses BERT as a cross-encoder to predict each pair of statements. This is not suitable for semantic search because there are too many pairs of combinations to calculate. SBERT is a modified BERT for Semantic Text Similarity (STS) work and is intended to address this issue. Unlike BERT, SBERT converts a statement to a fixed-length integration by applying a pooling layer to the output of the last level. The embedding of search queries and documents is calculated independently, allowing semantic searches using simple cosine similarities.

In our study, we use existing SBERT models for the semantic search of clinical descriptions as online documents for search queries. A pre-trained binary encoder is used to encode each article for semantic search. Each search query is encoded as a sentence embedding and related documents are searched. For each query, the 10 most relevant documents are retrieved and ranked based on the similarity of their sentence embedding and search query sentence embedding. After that, a pre-trained cross-encoder reorders the results with better quality. The utility function that measures whether a

document is relevant is based on whether the document mentions the disease names associated with the clinical description.

3.3.2.7 Evaluation Metrics

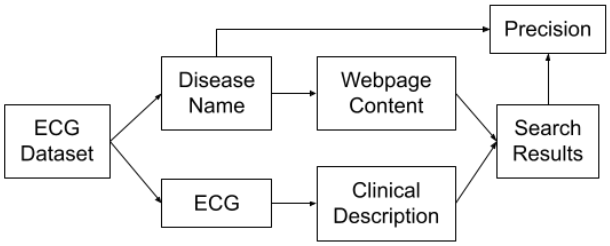

Figure 3.20 ***Process of Result Evaluation.***

Figure 3.20 shows how we evaluated our results. In the PTBDB dataset, ECGs were labelled with disease names. The articles of each disease name were collected. The clinical descriptions generated from ECG were used to rank the collected articles. The articles ranked in top 1, top 3, top 5 and top 10 were retrieved as search results. After that, a utility function was used to evaluate the retrieved article. The utility function measures whether a document is relevant based on whether that document refers to the patient's disease name. Finally, the precision of the search result was calculated as (3.21).

$$Precision = \frac{TP}{TP+FP} = \frac{TP}{S}, \tag{3.21}$$

where TP denotes the number of retrieved documents that are correct, FP denotes the number of irreverent retrieved documents, and S denotes the total number of search results.

The recall rate measures the percentage of retrieved articles taken from the *n*-ranking of all correct articles. Since there were more than 10 articles related to each disease, we only calculated the recall rate of the top 10 search results. The recall rate is calculated as (3.22).

$$Recall = \frac{TP}{TP+FN} = \frac{TP}{N}, \tag{3.22}$$

where TP denotes the number of retrieved documents that are correct, FN denotes the number of reverent documents that failed to be retrieved, and N denotes the total number of reverent documents.

3.3.3 Results

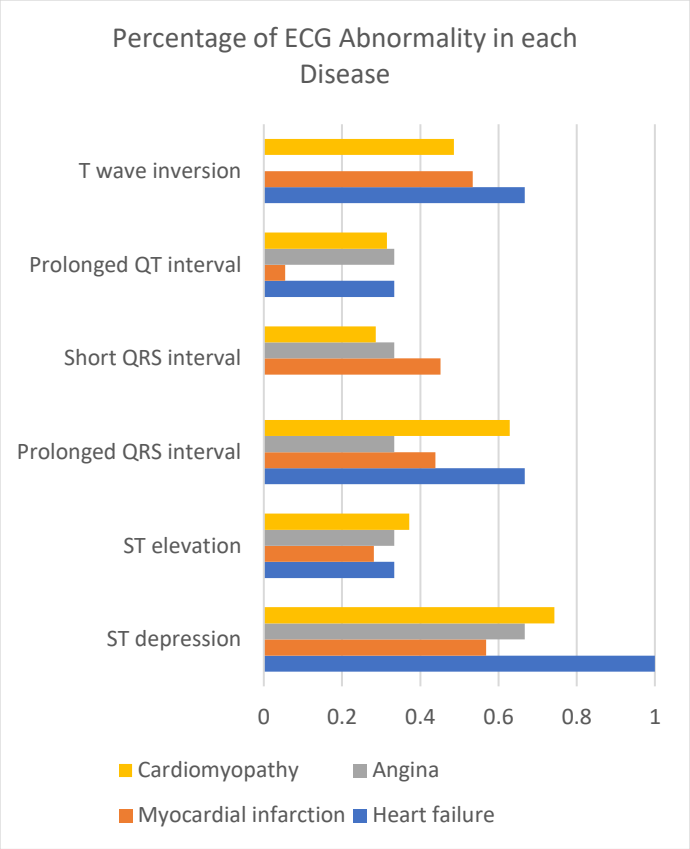

Figure 3.21 ***Percentage of ECG abnormality in each disease.***

Figure 3.21 shows the percentage of ECG abnormality detected by CardioVec for each disease. We found that a change in T wave could detect most cardiomyopathy, myocardial infarction and heart failure. T wave inversion and ST depression usually indicate heart failure. From Figure 3.21, we found that all the ECG abnormalities detected by CardioVec had crucial information for further diagnosis and information retrieval of each disease.

Table 3.14 ***Precision of Predicting Diseases.***

Name of Disease	Top 1 Precision	Top 3 Precision	Top 5 Precision	Top 10 Precision
Total	79.14%	55.61%	48.56%	46.84%
angina	66.67%	77.78%	73.33%	56.67%
myocardial infarction	75.34%	51.83%	45.34%	46.23%
heart failure	66.67%	44.44%	53.33%	56.67%
cardiomyopathy	97.14%	70.48%	59.43%	47.71%

Table 3.14 shows the searching precision of CardioVec. The search results achieved more than 79.14% of the top one precision of all ECGs. CardioVec achieved 79% top-one precision for all ECG, 75% top-one precision for ECG with myocardial infarction, and 97% top-one precision for ECG with

cardiomyopathy. We noted that heart failure and angina were found in a small number of samples (five in total), and CardioVec still achieved a good performance in detecting those diseases with more than 66% precision for the top one search result. The reason for this performance is that instead of supervised training, CardioVec classified the disease based on physiological measurement and knowledge online.

Table 3.15 ***Recall rate of predicting diseases.***

Name of Disease	Recall Rate
cardiomyopathy	26.51%
myocardial infarction	51.37%
angina	40.48%
heart failure	31.48%

Table 3.15 shows the average recall rate of documents related to different heart diseases. The search results achieved a recall rate of more than 50% each when searching documents for ECG with myocardial infarction. In a real scenario, a recall rate of 50% means that half the potentially useful documents can be searched by using clinical descriptions. With the huge amounts of data available on HSNs and the internet, CardioVec can provide users with rich information about their health.

3.3.4 Conclusion

This section introduces CardioVec, a framework designed for searching and indexing ECG data in HSNs and solves research question $RQ_{1.3}$ and hypothesis $H_{1.3}$. This section addresses the challenge of enabling the semantic search of text-based content from HSNs using ECG data from the IoT in healthcare. The challenge involves developing methods to convert biosignal data into clinically interpretable descriptions that can be used as keywords to search for relevant health information within HSNs.

This study aims to convert ECG data into interpretable clinical measurements that describe physical measurement abnormalities, thus providing keywords for semantic search within HSNs. To achieve this, we propose CardioVec, a system that utilises statistical analysis of specific physiological measurements extracted from the ECG. These measurements are then processed to generate human-readable keywords that can be used for searching for relevant information or connecting with other users on HSNs.

CardioVec allows users to obtain useful information about their cardiovascular health by generating human-readable keywords from the ECG. These keywords are generated using an evidence-based approach and can be used to search for relevant articles on the internet. The clinical descriptions are encoded as vectors to facilitate searching for relevant information. Our proposed CardioVec system improves the effectiveness and usefulness for patients with cardiovascular diseases who wish to learn about their potential health risks. Moreover, the vast amount of data available in

HSNs can be used to train the model and predict additional conditions, allowing users to automatically receive more information based on biomarkers without the need for manual searches.

The framework developed in this study can be extended and customised for automatic early diagnosis and mass cardiovascular health monitoring in both hospital and home settings. This can be combined with visualisation of cardiovascular physiological measurements, which can help trained physicians and parents track the cardiovascular health of term infants.

The contribution of this research is the development of CardioVec, which enables semantic search in HSNs by utilising clinically interpretable descriptions derived from ECG data. This improves search efficiency and enhances the trustworthiness of the reasoning process.

One limitation of this study is that the clinical descriptions generated by CardioVec are not exhaustive. They do not include subjective patient descriptions and other anomalies that were not covered in the mapping function. Future research could explore other measures of cardiovascular physiology to enable more effective and usable cardiovascular HIoT applications.

Additionally, it is important to note that the search results are currently evaluated based on whether the retrieved articles mention the disease name. In a real-world scenario, users would require more information than just the disease name. Further tests are needed to evaluate the performance of the search results and ensure their usefulness in practical applications.

4 Integrated HSN for Respiratory Diseases

4.1 Remote Diagnosis of Respiratory Conditions

4.1.1 Introduction

Acute respiratory infections, including pneumonia, are a major contributor to child mortality worldwide, particularly in developing countries (Ruuskanen, Lahti, Jennings, & Murdoch, 2011). In rural and remote areas, the challenges of a high patient-per-doctor ratio and limited access to advanced medical equipment further compound the difficulties in managing these respiratory diseases (B. Lei et al., 2014; Puustjärvi & Puustjärvi, 2011; I. Song, 2015). The use of IoT technology has the potential to overcome these challenges by enabling remote monitoring and diagnosis of respiratory health, providing valuable health-related information to patients (Azariadi et al., 2016). Therefore, remote diagnosis using Health IoT plays a critical role in respiratory disease management, particularly in remote areas. It not only reduces the need for unnecessary emergency department admissions but also has the potential to improve patient outcomes.

Table 4.1 ***State-of-The-Art of Diagnosing biosignal with Deep Learning.***

Application	**Date Type**	**Approach**	**Performance**
Classification of wheeze and crackle (Acharya & Basu, 2020)	Respiratory sound	CRNN	57.91% Macro F1-score, 66.38% Micro F1-Score
COVID-19 detection (Imran et al., 2020)	Cough sound	CNN	89.52% F1-score
COVID-19 detection (Laguarta et al., 2020)	Cough sound	CNN	0.95 AUC

Deep learning approaches provide an end-to-end solution for breath sound diagnosis. Table 4.1 illustrates the state-of-the-art in diagnosing breath sounds using deep learning. CNNs are particularly efficient for signal processing tasks due to their ability to leverage local relationship prior knowledge and compose lower-level elements into higher-level entities. CNNs have been widely employed in respiratory sound classification tasks, including wheeze and crackle detection (Acharya & Basu, 2020), as well as COVID-19 diagnosis (Imran et al., 2020; Laguarta et al., 2020; Subirana et al., 2020). On the other hand, LSTM utilises a memory block structure to establish long-distance relations, making them the most common RNN approach. In a study by Acharya and Basu (2020), a respiratory sound classification model for normal, wheeze, crackle, and combined (wheeze and crackle) sounds were proposed using CRNN-based on LSTM and CNN.

However, conventional deep learning methods, such as CNN and RNN, have limitations in modelling varying spatial relations and long sequences due to their inability to find relations with

long or varying distances (Vaswani et al., 2017). RNN possesses certain limitations when it comes to effectively modelling diverse spatial relations and handling long sequences, primarily due to issues such as gradient exploding or vanishing. CNNs encounter limitations when dealing with features that have varying distances, as they rely on fixed aggregation weights for spatially neighbouring input features (Hu et al., 2019).

Transformers excel in establishing connections between all nodes in an input sequence through a fully connected graph, facilitating the representation of flexible spatial and temporal distributions (Vaswani et al., 2017). This capability arises from the attention mechanism employed by Transformers, which learns the relationships among input timesteps independently of their positions. Consequently, Transformers exhibit adaptability to diverse feature distributions and mitigate the risk of overfitting.

To enhance the performance of the vanilla Transformer and its attention mechanism, researchers have proposed ALBERT and UT (Dehghani et al., 2018; Lan et al., 2019). ALBERT introduces parameter reduction techniques to enable faster training and improved utilisation of computational resources. On the other hand, the Universal Transformer extends the standard Transformer model by allowing dynamic computation during both the encoding and decoding stages, leading to enhanced representational power and higher accuracy. These advancements in Transformer-based models have demonstrated improved performance in various domains, such as NLP, document classification and machine translation.

In contrast, local relation approaches adopt a different strategy by calculating attention within a specific aggregation window, capturing local attention energy among inputs. This approach introduces an information bottleneck similar to CNN, which has proven effective for feature learning in CV tasks (Hu et al., 2019). LR-Net replaces the convolutional layers in ResNet with local relation layers, enabling the extraction of local relations within the image. Building upon LR-Net, the Swin Transformer has been proposed and has achieved state-of-the-art performance in image classification tasks (Liu et al., 2021). Its innovative architecture and attention mechanisms have contributed to significant advancements in the field of CV.

In addition, previous deep learning approaches for respiratory diagnosis were based on lung sounds collected with a microscope attached to the chest, which is not suitable for diagnosing children due to the requirement of trained medical professionals and the need for immobility during diagnosis (B. Lei et al., 2014; I. Song, 2015). Compared with lung sounds, breath sounds can be easily collected from the mouth and nose with a handheld microphone, making them suitable for non-invasive and contactless diagnosis (Lei, 2014; Song, 2015).

This research aims to develop a novel deep-learning algorithm for the non-invasive diagnosis of specific respiratory diseases, such as asthma or pneumonia, using breath sounds collected with mobile devices. The objective is to develop a deep learning algorithm that combines the advantages of CNN,

RNN and Transformer, to improve the accuracy of respiratory disease diagnosis, mitigate overfitting and overcome the limitations of current approaches.

This study proposes a novel deep learning algorithm called Recurrent Local Attention Transformer Encoder Classifier (ReLATEC), which utilises relative positional encoding, multi-head local attentions and recurrent encoder blocks for non-invasive diagnosis of respiratory diseases using sounds collected with mobile devices. In this study, we evaluate ReLATEC by comparing it with conventional CRNN and Transformer models using 1,144 breathing sounds recorded remotely using contactless non-invasive methods. ReLATEC achieved an AUC of 81.9% in accurately detecting the presence of respiratory health conditions in patients. Moreover, ReLATEC achieved AUC values of 0.77, 0.82, 0.83, and 0.68 for accurately detecting specific disease types: cold, flu, pneumonia and bronchitis, respectively. ReLATEC demonstrated a sensitivity of 86.98% and a specificity of 70.41% in predicting pneumonia, and a sensitivity of 80.38% and a specificity of 70.57% in predicting flu. The results achieved in flu and pneumonia detection have the potential to contribute to the effective control of infectious respiratory diseases. In comparison to CRNN, ReLATEC effectively mitigates overfitting and yields robust results for the diagnosis of respiratory health conditions based on biosignals.

To our knowledge, this is the first time that relative positional encoding, multi-head local attention and recurrent encoder block have been combined. In doing so, it has reduced overfitting, memory usage and training time, leading to better overall performance by 5.2% over the Transformer in detecting respiratory problems using breath sounds collected remotely. The results show that respiratory health conditions including pneumonia can be detected remotely, non-invasively. and at a low cost. It further shows that respiratory health diagnosis can be done using breathing sounds collected from the mouth and nose through any mobile phone.

The contribution of this study is introducing ReLATEC as a novel deep learning algorithm specifically designed for non-invasive diagnosis of respiratory diseases using breath sounds collected with mobile devices. The algorithm combines the advantages of CNN, RNN and Transformer models, leveraging their strengths in feature composition, temporal modelling, and flexible spatial relations. The evaluation of ReLATEC against CRNN and vanilla Transformer demonstrates its superior performance and reduced overfitting, highlighting its potential for accurate and efficient respiratory disease diagnosis.

4.1.2 Method

4.1.2.1 Dataset

Table 4.2 ***Data Source.***

Dataset	Cold	Flu	Pneumonia	Bronchitis	Healthy	Total
DK1	40	15	29	32	61	130

DK2	342	341	23	322	172	653
NS	334	7	182	167	21	361
Villages	36	21	0	0	346	403
Total	752	384	234	521	600	1144
Total %	48.61%	24.82%	15.13%	33.68%	38.78%	100%

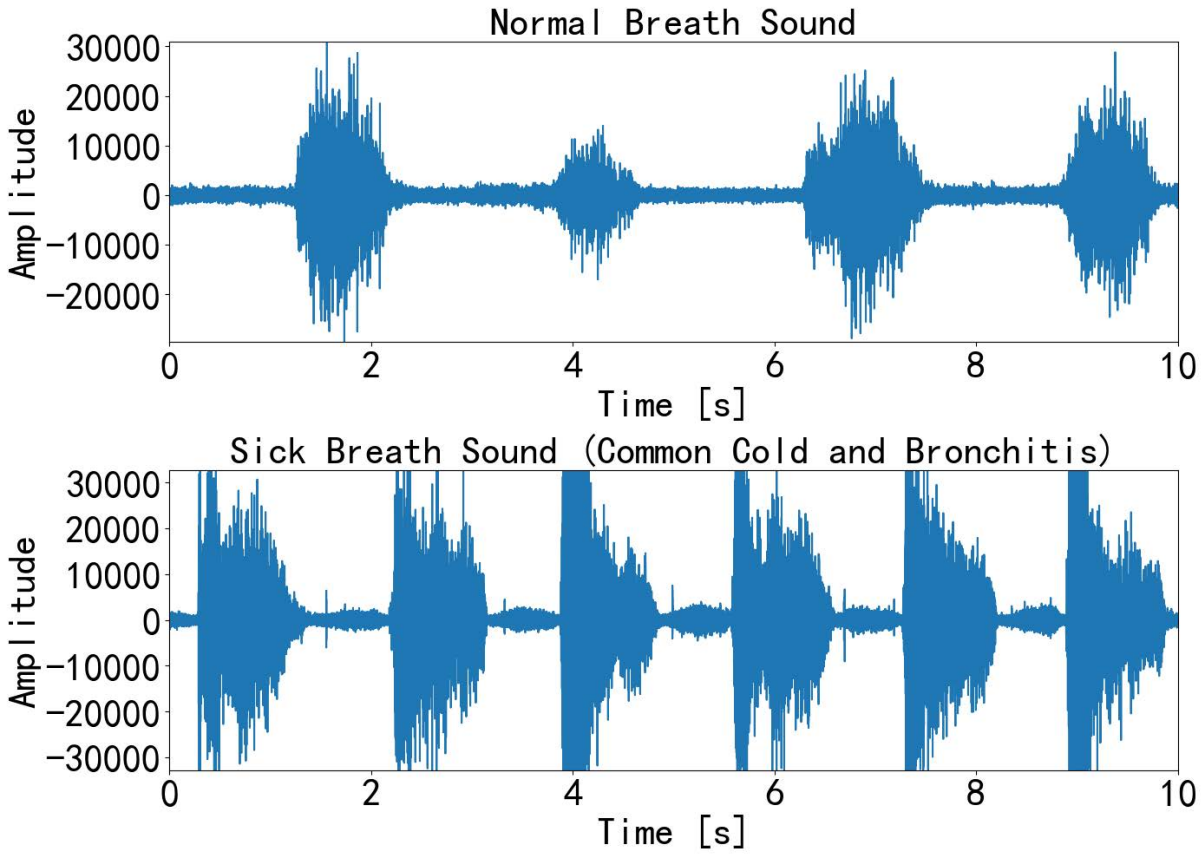

Figure 4.1 ***Examples of normal breath sound and sick breath sound.***

The dataset that we used comprised breath sounds from 1,144 patients in Bangladesh as part of Bill Gates' Global Health Challenge initiative project. The sounds were collected from infants and child patients (aged 0.6 months to 7 years) having one or more of the following respiratory health conditions: pneumonia, bronchitis, cold and flu (Table 4.2). Figure 4.1 displays two examples from our dataset. The upper plot represents a healthy breath sound, while the lower plot depicts an example of a breath sound associated with common cold and bronchitis. In this example, breath sounds with common cold, and bronchitis are more frequent and deeper in nature. The sounds were collected using microphones placed near the nose or mouth of patients without touching. DK1 and DK2 datasets were collected from two separate hospitals in Dhaka. NS dataset was collected from a hospital in Narsingdi. DK1, DK2, and NS data are collected using digital sound recorders (SONY ICD-UX523F). The village dataset was collected from the Bhola district as a pilot test data using Nokia cell phones and mobile messages.

The sampling rate of clinical data (DK1, DK2, NS) collected at hospitals was 44.1kHz (I. Song, 2015), whereas the sampling rate of the pilot study data (Village) was 22kHz. We labelled the collected sounds with a specific diagnosis of the patients by trained doctors at the hospitals. We also labelled the starts and ends of exhales and inhales of the breathing cycles.

To test whether our prediction models would work on the sounds collected at different recording conditions, we resampled the original sound data at four different sampling rates (8kHz, 11.025kHz, 16kHz and 22.05kHz). The resampled sound data were then resampled to 44.1kHz to generate standard prediction models that can be used for breath sounds collected at any sampling rate.

Table 4.3 ***Total numbers of segments divided dataset to rebalanced data.***

Dataset	Cold	Flu	Pneumonia	Bronchitis	Disease	Healthy	Total
Test	691	316	169	417	790	442	1232
Training	2241	1114	778	1654	2791	1437	4408
Validation	232	117	68	169	307	183	490
Total	3164	1547	1015	2240	3888	2242	6130
Total %	51.6%	25.2%	16.6%	36.5%	63.4%	36.6%	100.0%

The dataset was divided into three sets: a training set, a validation set, and a testing set, each serving different purposes. To ensure the integrity of the data, we allocated 20% of the individual records per patient to the test dataset. The remaining records were then split into the training set and validation set, with 80% assigned to the training set and 20% to the validation set. This division guarantees that no patient data appears in both the training and validation sets, as well as the test set.

We selected multiple sound segments with a duration of 10 seconds for each breath sound, ensuring that each segment contained at least three breathing cycles. Table 4.3 provides the final count of sound segments for patients with cold, flu, pneumonia, bronchitis, healthy individuals, and those categorised as “sick” (meaning they had at least one of the four conditions) for data argumentation. It is worth noting that some patients may have multiple labels, which is why the total number of segments for the four disease categories may exceed the number for the sick or total patient groups in each dataset. To extract features, we generated the power spectrogram of the sound, comprising 128 channels with a window size of 50ms and a window step of 20ms.

4.1.2.2 Recurrent Local Relation Transformer Encoder Classifier (ReLATEC)

Deep learning approaches provide an end-to-end solution for analysing breath sound: including feature selection feature engineering, and prediction. CNN and RNN were the most common deep-learning models for biosignal processing tasks. CNN has been widely used in respiratory sound classification tasks, such as wheeze and crackle detection (Acharya & Basu, 2020) and COVID-19 diagnosis(Imran et al., 2020; Laguarta et al., 2020; Subirana et al., 2020).

However, conventional deep learning methods, such as CNN and RNN, have limitations in modelling varying spatial relations and long sequences due to their inability to find relations with long or varying distances (Vaswani et al., 2017). To address these limitations, the Transformer model was introduced, which can relate all nodes in the input sequence with a fully connected graph, enabling flexible spatial and temporal distributions (Vaswani et al., 2017). The Transformer's attention mechanism learns the relations among input timesteps independently of their positions, making it adaptable to different feature distributions and preventing overfitting. LR-Net (Hu et al., 2019), ALBERT (Lan et al., 2019), and UT (Dehghani et al., 2018) have been further proposed to improve upon the vanilla Transformer and achieve better performance.

Table 4.4 ***Comparison of different models.***

Model	**Relation Type**	**Relation Range**	**Cross-Layer weight sharing**	**Risk to gradient vanish or exploding**
ReLATEC	**Dynamic**	**Local**	**Yes**	**Low**
ALBERT / UT (Dehghani et al., 2018; Lan et al., 2019)	**Dynamic**	Global	**Yes**	**Low**
LR-Net (Hu et al., 2019)	**Dynamic**	**Local**	No	**Low**
Transformer (Vaswani et al., 2017)	**Dynamic**	Global	No	**Low**
CNN	Static	**Local**	No	**Low**
RNN	Dynamic	Global	No	High
CRNN	Mixed	Mixed	No	Medium

Therefore, for diagnosing respiratory diseases, we developed a new deep-learning algorithm called ReLATEC that utilises a customised Transformer Encoder network. This study proposes a novel deep learning algorithm named Recurrent Local Relation Transformer Encoder Classifier (ReLATEC) that utilises relative positional encoding, multi-head local attentions and recurrent encoder blocks to diagnose respiratory diseases non-invasively using sounds collected with mobile devices. We evaluate the performance of ReLATEC against LR-Net, UT, CRNN and vanilla Transformer and show that ReLATEC outperforms these methods, demonstrating better performance and reduced overfitting.

Table 4.4 compares the key features of ReLATEC, UT, Transformer and the other conventional deep learning approaches. ReLATEC combines the local relation from LR-Net and the Recurrent encoder block from UT. The combination of the methods showed significantly better performance and reduced overfitting compared with CRNN and vanilla Transformer in our evaluations. Furthermore, ReLATEC uses relative positional encoding instead of global positional encoding to remove the absolute positional information, which is not meaningful in modelling biosignals.

Transformer significantly improves the performance on NLP, audio and CV tasks. However, to our best knowledge, the Transformer utilising both recurrent structures with local attention has not yet been used for diagnosing breath sound. Therefore, we propose ReLATEC that combines recurrent encoder block and local relation.

ReLATEC is a modified Transformer for biosignal processing to replace CNN or RNN, based on the evidence of high performance and the potential of different types of self-attention-based approaches. Furthermore, the ideas of cross-layer weight sharing from ALBERT and windowed attention from LR-Net are combined in ReLATEC.

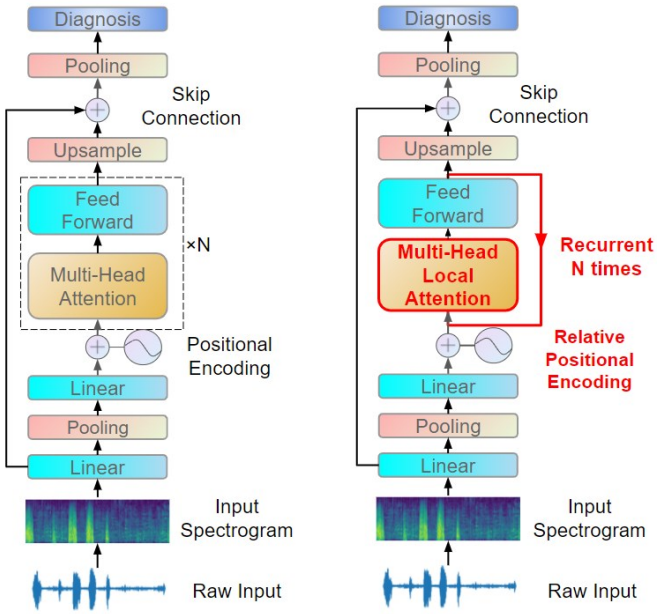

Figure 4.2 ***Comparison of vanilla Transformer (a) and ReLATEC (b).***

Figure 4.2 compares the procedures of ReLATEC and Transformer. The difference between ReLATEC and Transformer is highlighted with bold red font. Firstly, ReLATEC uses relative positional encoding instead of global positional encoding. Secondly, ReLATEC uses a local relation instead of global relations to generate contextualised features using a local aggregation window of the input. Thirdly, ReLATEC uses a recurrent encoder block instead of multiple individual encoder blocks to reduce model parameters.

The entrance blocks of ReLATEC and Transformer are the same. Raw input is first converted into a spectrogram with the shape of $N \times D$. N is the spectrogram frame number and D is the channel number. The spectrograms are then used as input feature maps of Transformer or ReLATEC. After that, a linear layer is used as an embedding layer to convert the frames into embeddings $N \times D$, which is like converting tokens into embeddings. The embeddings are then input to the Transformer blocks or the ReLATEC block. The outputs of the Transformer blocks and the ReLATEC block have the same shape as $N \times D$. Finally, a global pooling layer reduces the output of the model to the shape of $1 \times D$.

ReLATEC combines the advantages of local relations and recurrent structure. Local relations focus on contextual features in local scales to improve the performance in audio-related tasks. The introduction of the recurrent structure is equivalent to cross-layer weight sharing, which minimises the parameter while improving its performance.

To process audio signals with ReLATEC, we first convert time series biosignal data into 2D-spectrograms. The spectrogram is then input into the entry block as (4.1).

$$E = f\left(n_2\left(W_2 Pool2\left(f\left(n_1(W_1 X + b_1)\right)\right) + b_2\right)\right) + P, \tag{4.1}$$

where *X* denotes the input; *W*, *b* denotes the weight and bias of each layer; *P* denotes positional encoding; *f* is Gaussian Error Linear Units (GeLU) as the activation function; *n* is layer normalisation function; *Pool2* is the pooling function with window size and stride size of two. In the entry block, a linear layer extracts features for each time step from the spectrogram. A max-pooling layer with a size of two is used to down sample the time sequence. The down sample features are further converted into embedding vectors. To represent position information, positional embedding is added to the embedding sequences. After that, output *E* feeds into the ReLATEC layer six times recursively. Finally, the last output of the ReLATEC transforms into a vector with a dimension of output.

ReLATEC learns local relations of features rather than global relations of features as in Transformer. The local relation in ReLATEC is constraining the aggregation window of self-attention.

We hypothesise that using windowed attention is crucial for feature learning in the audio analysis resulting in improved performance with less risk of overfitting.

ReLATEC learns local relations of features rather than global relations of features as in Transformer. The local relation in ReLATEC is constraining the aggregation window of self-attention. We hypothesise that using windowed attention is crucial for feature learning in audio analysis resulting in improved performance with less risk of overfitting.

In windowed self-attention, only the time steps inside the sliding aggregation window are used to calculate the attention energy. The localised attention function *LA* shown below replaces the attention function defined in (4.2-4.3):

$$LA(Q, K, V) = mask\left(softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)\right)V, \tag{4.2}$$

$$mask(E_{i,j}) = \{E_{i,j}, |i - j| < \frac{windowsize}{2} + 1\ 0, otherwise\ , \tag{4.3}$$

where Q, K and V are query, key and value, respectively; d_k is the dimension of the key, i, j is the index of query and key; E is attention energy. In our approach, we test the length of the sliding window of 11 and 15.

To represent positional information in Transformer, positional embedding is appended to the output of the entry block as illustrated in Equation (4.1) above. The positional encoding is calculated as (4.4) and (4.5):

$$(p, 2i) = sin\left(\frac{p}{b^{\frac{2i}{d_k}}}\right), \tag{4.4}$$

$$(p, 2i+1) = cos\left(\frac{p}{b^{\frac{2i}{d_k}}}\right), \tag{4.5}$$

where p is the position, b is a constant, i is the dimension number, and d_k is the total dimension number of the key. In the original Transformer approach, the base b is set to 10000. However, the positional encoding of the vanilla Transformer contains absolute position information.

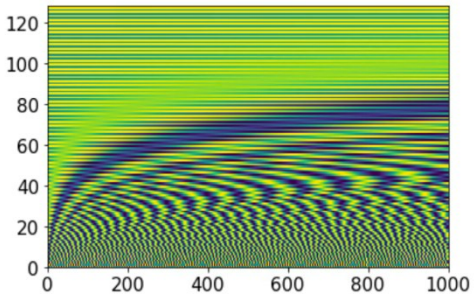

Figure 4.3 ***Conventional local positional encoding in Transformer.***

Figure 4.3 shows the positional encoding when the sequence is long. When the time step is below 200, the encodings for those positions are very different and may cause overfitting. The higher dimension contains absolute position information since it does not repeat its pattern periodically. On the other hand, the lower dimension has a repeating pattern. The complete information is an interference in our approach since breath sound is not a sentence and only the local position information is useful. Thus, those absolute positional encoding leads to overfitting.

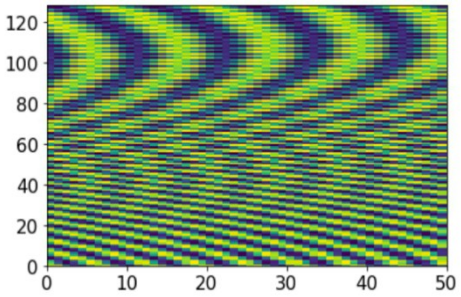

Figure 4.4 ***Customised local positional encoding in ReLATEC.***

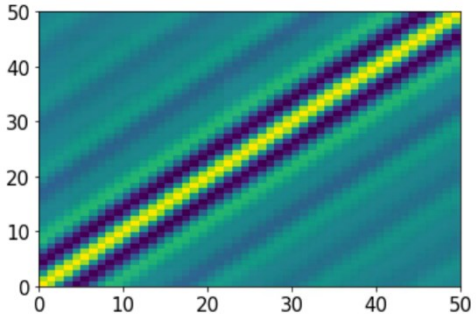

Figure 4.5 ***Dot-product between encoding in different positions.***

Therefore, we used a novel local positional encoding that contains local information to prevent the overfitting problem caused by absolute position. To remove this absolute positional information, the base is set to 2. The value of the higher dimension in the positional encoding with the base of 2 is equal to the value of the lower dimension in the positional encoding with the base of 10000. Figure 4.4 shows the pattern of the modified positional encoding. The pattern is now repeating itself in a period for all dimensions. Figure 4.5 shows the dot-product between the encoding. With a window size of 11 for the windowed attention, only the dot-product between 5 is calculated. In this range, the dot-product similarity gradually decreases.

ReLATEC has only one Transformer Encoder block which is iterated six times. This recurrent structure is based on previous approaches such as UT {Dehghani, 2018 #359} and ALBERT (Lan et al., 2019). UT and ALBERT outperform vanilla Transformers in different tasks.

Like ALBERT, ReLATEC has only one encoder block that computes contextual features iteratively. The following (4.6), (4.7), and (4.8) show the forward step of a ReLATEC block.

$$X_i = ReLATEC\ layer(X_{i-1}), \tag{4.6}$$

$$ReLATEC\ layer(X_i) = FFN(X_i + LA(X_i, X_i, X_i)), \tag{4.7}$$

$$FFN(X_i) = n_2(X_i + f(n_1(W_1(W_0 X_i + b_0))) + b_1), \tag{4.8}$$

where X_i is output from iteration i; W and b denote learnable parameters; n is layer normalisation function; f is GeLU activation function; *FFN* is a feed-forward network containing two linear layers; *LA* is multi-head localised self-attention in (6.2). The multi-head attention is used as self-attention, where query, key and value are all input. The multi-head self-attention uses local attention in equations (3) and (4).

4.1.2.3 Experiment Parameters

Table 4.5 ***Layer configuration of CRNN.***

Block	**Layer Type**	**Kernel**	**Output Channel**
Entry Block	Linear	GeLU	64

	Pooling 2	-	-
Residential Block x3	Residential Layers x3	GeLU	64
	Pooling 2	-	-
RNN	BiLSTM	-	128
	LSTM (last output)	-	128
Classification	Linear	GeLU	64
	Linear	Sigmoid	1

Table 4.6 ***Layer configuration of Transformer-based models.***

Block	**Layer Type**	**Kernel**	**Output Channel**
Entry Block	Linear	GeLU	64
	Pooling 2	-	-
	Linear	GeLU	128
	Positional Encoding		128
Transformer Encoder Block	Attention Layer	-	128
	Linear	GeLU	64
	Linear	Linear	128
Classification	Adaptive Pooling	-	-
	Linear	Sigmoid	1

We compared ReLATEC with other benchmark approaches for performance in binary classification tasks. To test the effectiveness of local relation and cross-layer weight sharing in ReLATEC, the benchmark criteria included other architectures of Transformer. The approaches were: ReLATEC (Transformer with local relation and cross-layer weight sharing); Recurrent Transformer (Transformer with cross-layer weight sharing); One Layer Transformer (Transformer with one Transformer Encoder block only); vanilla Transformer (original Transformer with six independent Transformer Encoder blocks) and CRNN. We tried two different window sizes of ReLATEC local attention. Table 4.5 and Table 4.6 show the layer configuration and structure of the CRNN and Transformer-based models. The Feed-Forward layer in the Transformer Encoder block had 32 hidden units to reduce overfitting by feature selection. The layer configuration of CRNN was based on a previous study (Acharya & Basu, 2020), which was also a respiratory sound classification approach and optimised with trial-and-error to achieve the best performances. The models were implemented with PyTorch and trained on Nvidia GeForce RTX 2080. We trained the ReLATEC model and other benchmark models using a training configuration of 50 epochs and batch sizes of 32. The models were trained using the AdamW (Adam with Weight Decay) optimizer with an initial learning rate of 0.0001 and L2 normalization factor of 0.005. The learning rate was adjusted using a step decay learning rate

scheduler, which reduced the learning rate by 20% after each epoch. Additionally, early stopping was implemented with a patience of 5 epochs to prevent overfitting.

We conducted experiments using CRNN, Transformer, Recurrent-Transformer, and ReLATEC models for both binary classification and multiple label classification tasks. For binary classification, the models were trained to distinguish between two outcomes: sick or healthy, based on the presence or absence of any medical condition. Cross entropy loss was utilized during the training process for binary classification. On the other hand, multi-class models were trained to handle the complexity of categorizing data into multiple distinct diseases. Binary Cross Entropy (BCE) was used for binary classification.

4.1.2.4 The Evaluation Matrices

The evaluation matrices of our experiments are Sensitivity, Specificity, Accuracy, Precision, and AUC of Receiver Operator Characteristics (ROC). Precision is calculated as:

$$Precision = \frac{TP}{Pred}, \tag{4.9}$$

where *TP* is the number of true positives; *Pred* is the number of cases predicted as positive cases.

AUC is calculated as the area of the ROC curve.

4.1.3 Result and Analysis

4.1.3.1 Single Label Classification

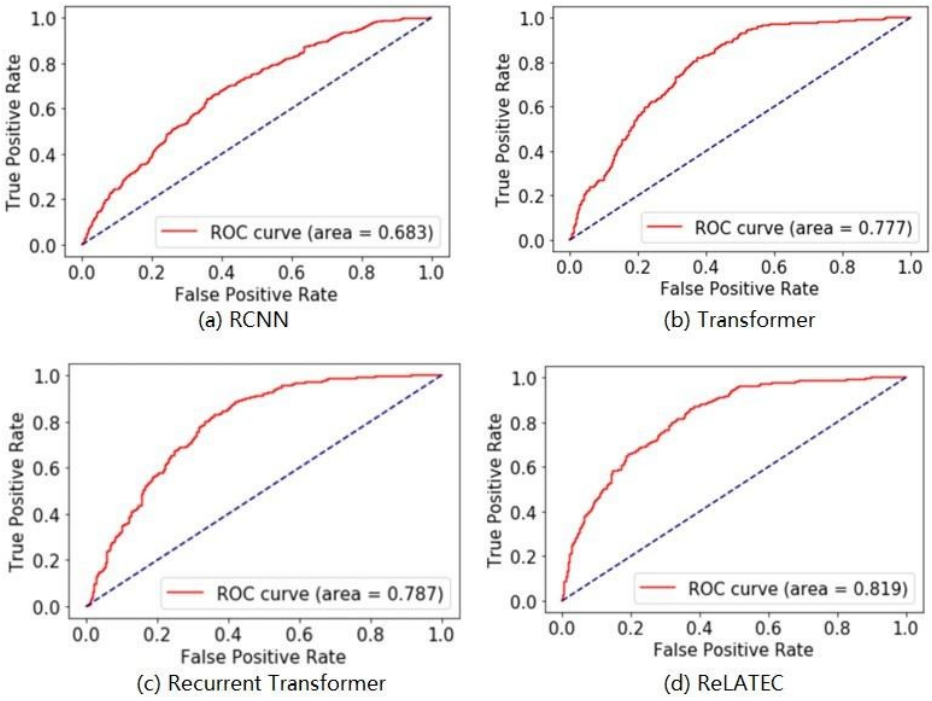

Figure 4.6 ***ROC of and ReLATEC (window size 11) for binary classification of disease from healthy.***

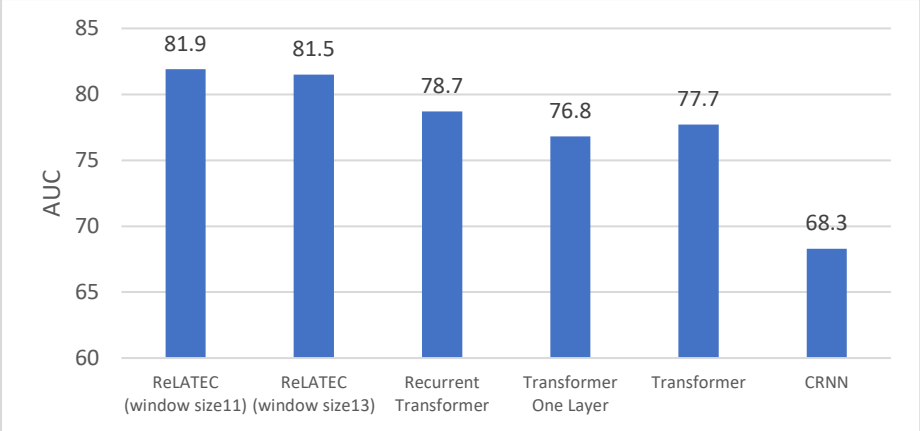

Figure 4.7 ***AUC of different approaches (in percentages).***

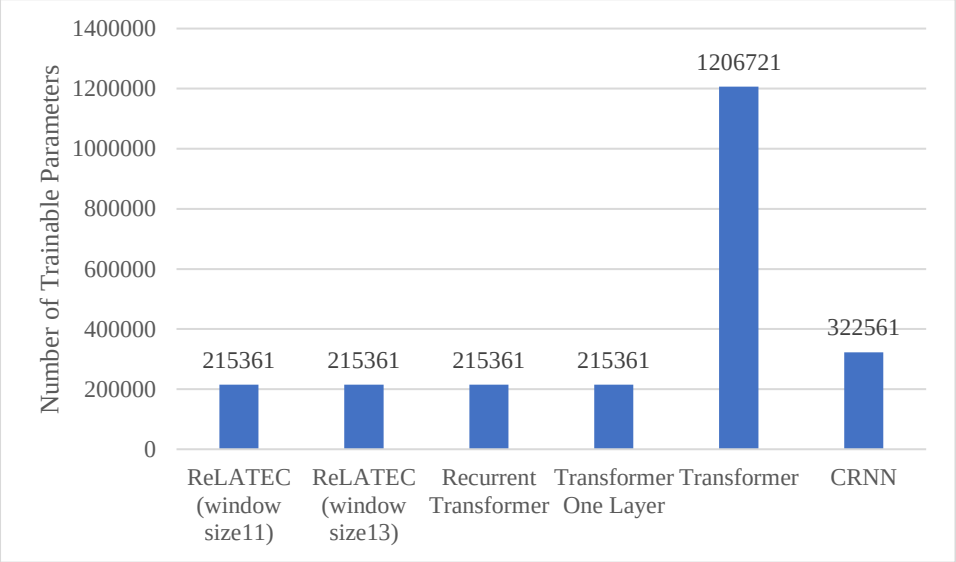

Figure 4.8 ***Number of trainable parameters of different approaches.***

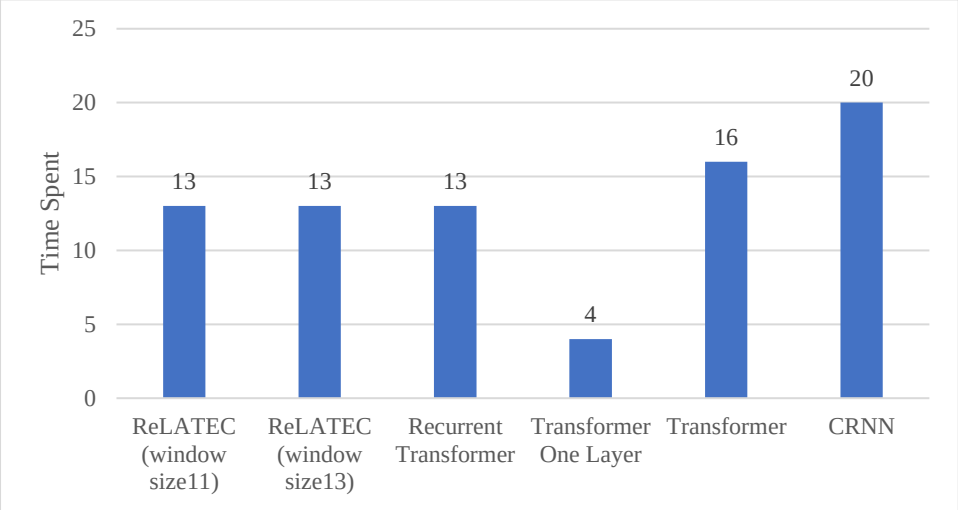

Figure 4.9 ***Time spent in seconds for all approaches.***

To validate the effectiveness of recurrent structure and local relation of ReLATEC, we compared different approaches in terms of classification performance, number of trainable parameters and time spent in training steps. Figure 4.6 illustrates ROC of and ReLATEC for binary classification of disease from healthy. Figure 4.7 compares the AUC values of ROC curves of ReLATEC and the other benchmark approaches. ReLATEC significantly outperformed CRNN with AUC values of 81.9% and 64.1% for ReLATEC and CRNN, respectively. Figure 4.8 compares the number of trainable parameters of different approaches. Transformers with one encoder block, including recurrent Transformer and ReLATEC, significantly reduce memory usage compared to vanilla Transformer and CRNN. Although vanilla Transformer significantly outperforms CRNN, it requires 1.206 million trainable parameters, while CRNN only requires 0.323 million parameters. On the other hand, ReLATEC only has 0.215 million parameters. Figure 4.9 compares the time spent on one single epoch of training different models. ReLATEC only requires around 13 seconds, while Transformer requires 16 seconds, and CRNN requires 20 seconds. ReLATEC and recurrent Transformer take less time for training. Compared to CRNN and vanilla Transformer, ReLATEC achieved better performance with fewer parameters and shorter training time.

We should note that the one-layer Transformer did not perform better than the recurrent transformer and the vanilla transformer indicating that multiple encoder blocks are necessary for the Transformer to achieve good performance, regardless of weight sharing. Furthermore, the recurrent Transformer outperformed the vanilla Transformer indicating that multiple encoder blocks with cross-layer weight sharing of recurrent Transformer and ReLATEC can result in improved performances. Recurrent Transformer Encoder significantly reduces the number of parameters without impacting the performance negatively. With a reduced number of parameters, the risk of overfitting is minimised.

ReLATEC is a further improvement of Recurrent Transformers by using local relations rather than global relations. Local relation improves the AUC by up to 81.9 with a window size of 11. These results show that constraining the aggregation window of self-attention improves performance. By combining recurrent structure and local relation, our proposed ReLATEC greatly improves the AUC over CRNN and vanilla Transformer.

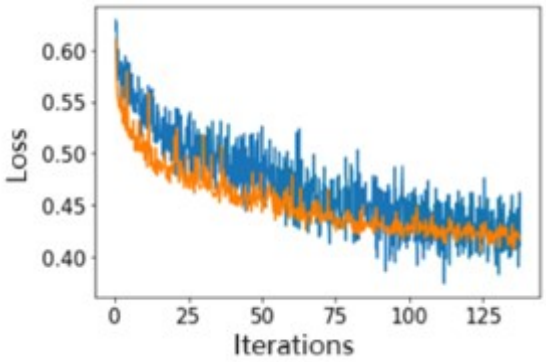

Figure 4.10 ***Training loss (blue) and validation loss (orange) of ReLATEC.***

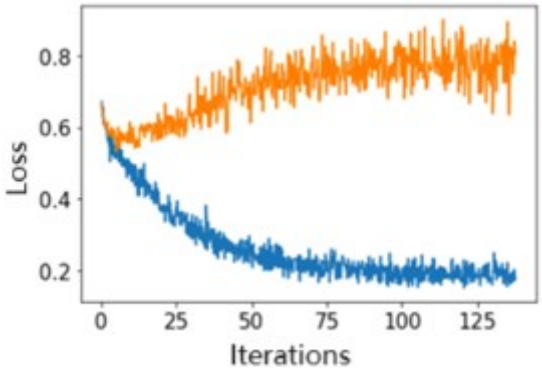

Figure 4.11 ***Training loss (blue) and validation loss (orange) of CRNN.***

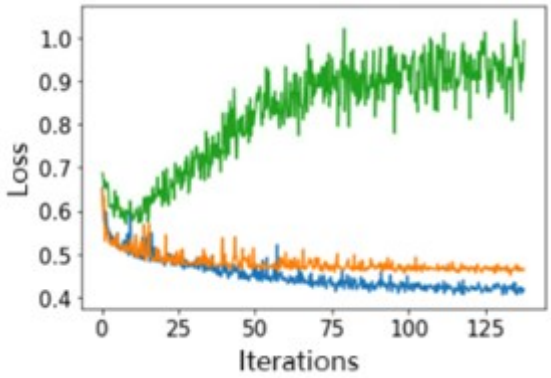

Figure 4.12 ***Validation loss of ReLATEC (blue), Transformer (orange), and CRNN (green).***

Now we look at how ReLATEC solves overfitting. Figures 4.10 and 4.11 show the training loss (blue) and validation loss (orange) of ReLATEC and CRNN. The training rate and L2 normalisation factor of both approaches are 0.0001 and 0.005, respectively. CRNN shows a serious overfitting problem after 10,000 iterations and eventually emerged unusable. ReLATEC and Transformer had a minimal overfitting problem, and validation loss did not increase. ReLATEC and Transformer did not memorise and specialise training data as CRNN does. Figure 4.12 compares the validation loss of the CRNN, vanilla Transformer and ReLATEC. Transformer-like approaches, including ReLATEC, significantly reduced overfitting. Using cross-weight sharing and local relation, ReLATEC further reduced the overfitting problem compared with the vanilla Transformer. ReLATEC has a slight advantage in overcoming the overfitting problem and converging speed. Although early stopping was applied in all approaches, ReLATEC outperforms CRNN and Transformer in validation loss. This difference in validation loss indicates that ReLATEC overcomes the overfitting problem as expected.

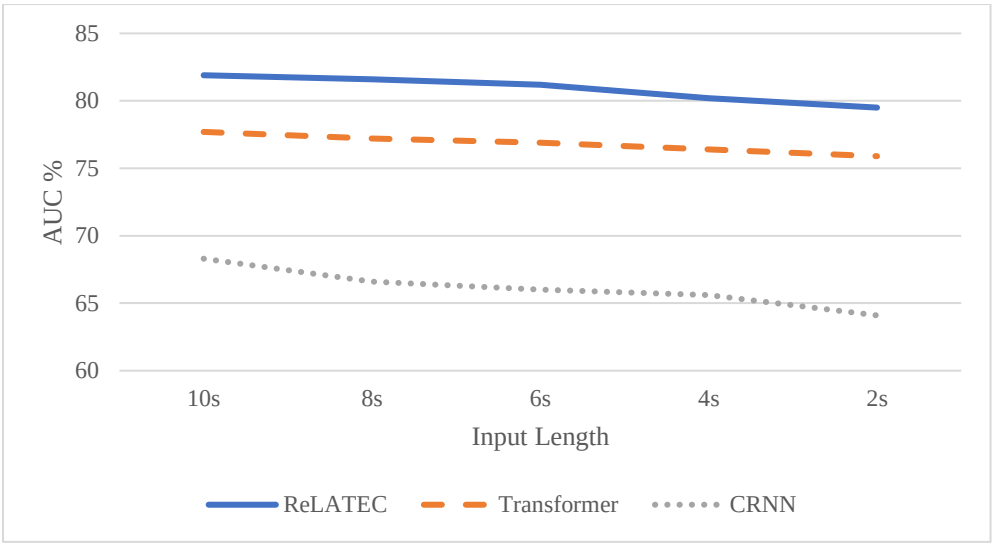

Figure 4.13 ***Line chart of AUC for ReLATEC, Transformer and CRNN with varying lengths of the input sound.***

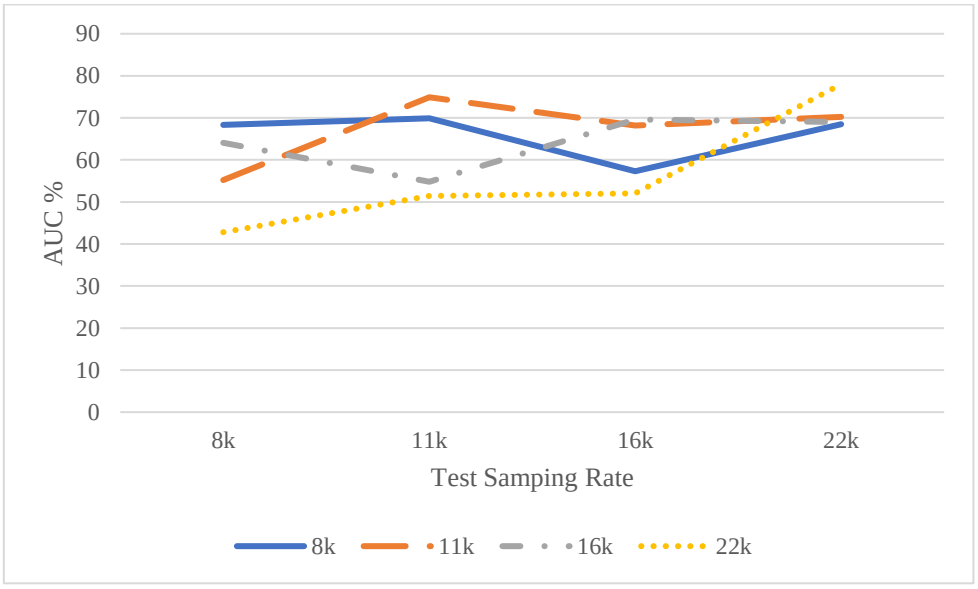

Figure 4.14 ***AUC of ReLATEC with varying test sampling rates.***

ReLATEC was stable with varying lengths and quality of audio input. Figure 4.13 shows the performance of the classifiers when the length of the input sound is shorter than the required 10 seconds, as the models are trained with 10-second length segments. Both Transformer and ReLATEC outperformed CRNN. Even with two seconds of sound, ReLATEC achieved the AUC of 79.5 demonstrating its robustness to input data quality. Figure 4.14 shows the AUC of ReLATEC with varying test sampling rates. ReLATEC was more robust to different sound recording conditions than other approaches. It provided robust performance across the input sounds recorded at different simulated sound sampling rates.

Table 4.7 ***Performance with over 95% sensitivity using 8k sampling rate.***

Model	**AUC**	**Sensitivity**	**Accuracy**	**Precision**	**Specificity**
ReLATEC	81.9%	95%	72.3%	65.3%	49.5%

Transformer	77.7%	95%	70.4%	63.5%	45.0%
CRNN	68.3%	95%	57.6%	54.3%	19.91%

Table 4.7 shows the cost-saving by using a ReLATEC classifier, which was configured to achieve 95% sensitivity (recall rate). Given an infection rate of 10%, ReLATEC reduces painful and costly RT-PCR-based tests by 44.5% since the specificity is 49.5%. Only 49.5% of the 90% healthy population needs further testing.

Table 4.8 ***Performance with over 95% sensitivity using 8k sampling rate.***

Application	**Date Type**	**Feature**	**Approach**	**Performance**
Classification of wheeze and crackle (Acharya & Basu, 2020)	Respiratory sound	Mel-spectrogram	CRNN	57.91% Macro F1-score, 66.38% Micro F1-Score
COVID-19 detection (Imran et al., 2020)	Cough sound	MFCC	CNN	92.64% Accuracy
COVID-19 detection (Laguarta et al., 2020)	Cough sound	MFCC	CNN	95.0% AUC
Respiratory insufficiency detection in COVID-19 (Gauy & Finger, 2021)	Speech	MFCC	Transformer	96.38% Accuracy
Classification of wheeze and crackle (Moummad & Farrugia, 2022)	Lung Sound	Mel-spectrogram	CNN	57.55 Score (Means of sensitivity and specificity)
Diagnosis of respiratory diseases (Basu & Rana, 2020)	Lung Sound	MFCC	RNN (GRU)	95.67%
Ours	Breath sound	Mel-spectrogram	CRNN	64.1% AUC
Ours	Breath sound	Mel-spectrogram	ReLATEC	81.9% AUC

Table 4.8 presents the state-of-the-art in diagnosing respiratory sounds using deep learning techniques. CNN has been employed for the diagnosis of COVID-19 using cough sounds (Imran et al., 2020; Laguarta et al., 2020). Acharya and Basu (2020) proposed a respiratory sound classification method using CRNN, a combination of CNN and RNN, to classify normal, wheeze, crackle and both wheeze and crackle sounds. Transformers have also been utilised for the diagnosis of COVID-19 using speech sounds, achieving an accuracy rate of 96.38%. The most common features in deep learning-based approaches are the Mel-spectrogram and MFCC. State-of-the-art approaches involve

the use of Transformers and CNN on lung sounds, and we conducted a comparative analysis using these approaches on our dataset. Currently, there is no existing approach that specifically diagnoses respiratory diseases in children using non-contact breath sounds. The CRNN-based method demonstrates lower performance compared with the state-of-the-art approaches, as indicated by lower accuracy rates or other relevant metrics. The quality of the signal and the characteristics of the population group pose barriers to diagnosis performance. In comparison to the existing approaches, our proposed ReLATEC method is non-invasive and achieves similar performance to the state-of-the-art approaches.

4.1.3.2 Multiple Label Classification

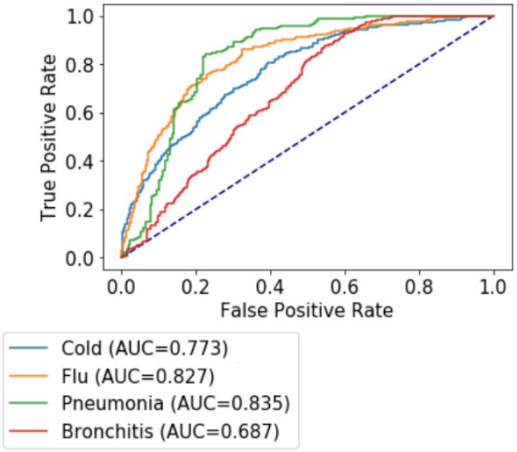

Figure 4.15 ***ROC of cold, flu, pneumonia and bronchitis prediction.***

Table 4.9 ***Performance of multiple Label Classification.***

Model	AUC	Accuracy	Precision	Sensitivity	Specificity
cold	77.33%	69.41%	69.66%	68.76%	70.06%
flu	82.74%	75.47%	73.20%	80.38%	70.57%
pneumonia	83.48%	78.70%	74.62%	86.98%	70.41%
bronchitis	68.68%	60.79%	63.24%	51.56%	70.02%

In this section, we present the performance of ReLATEC in conducting multiple label classification for predicting various diseases. Figure 4.15 displays the ROC curves for each disease classification, while Table 4.9 provides information on AUC, accuracy, precision, sensitivity and specificity for each disease.

The outcomes for flu and pneumonia prediction were exceptional, with both achieving an AUC of more than 80% and sensitivity exceeding 80%. Notably, the model demonstrated a sensitivity of 86.98% and a specificity of 70.41% for pneumonia prediction. These results highlight the cost efficiency and effectiveness of our model in predicting pneumonia and flu from breath sounds. The

model's performance has the potential to assist patients in confirming their health concerns promptly and can potentially save up to 70% of unnecessary costs.

4.1.4 Discussion

The proposed ReLATEC utilises local attention and recurrent structure on the Transformer framework for sound-based diagnosis tasks, demonstrating more robust results compared with conventional approaches. Local attention calculates attention energy within a fixed-sized window, resulting in linear spatial complexity, in contrast to the original Transformer's O(n^2) spatial complexity. This combination of methods has shown significantly improved performance, reduced memory usage, training time and overfitting compared with CRNN, Recurrent Transformer and vanilla Transformer models. However, further work is still needed to test ReLATEC on a larger dataset and deploy this study for real-world applications, specifying the specific objectives or goals of this deployment.

The proposed ReLATEC, with its combination of local attention and recurrent structure, is to significantly improve the performance of diagnosing respiratory diseases using breath sound, outperforming the vanilla Transformer and conventional CRNN models. ReLATEC has exhibited relatively good performance for infectious diseases such as flu and pneumonia. Furthermore, since breath sound collection only requires a mobile phone and does not necessitate well-trained medical professionals, our proposed approach shows the potential to provide fast, non-invasive and low-cost pervasive respiratory health screening for children.

4.1.5 Conclusion

This section introduces a novel deep learning algorithm named ReLATEC, which addresses research question $RQ_{2.1}$ and hypothesis $H_{2.1}$. The research question and hypothesis being addressed by ReLATEC should be briefly explained for better clarity. The section focuses on the challenge of developing an efficient deep-learning model that effectively combines CNNs, RNNs and Transformers for the analysis of breath sounds. Specifically, RNNs face limitations in modelling diverse spatial relations and long sequences due to issues of gradient exploding or vanishing. In contrast, CNNs excel at composing lower-level elements into higher-level entities, making them well-suited for signal processing tasks. However, CNNs struggle to model features with varying distances due to the fixed aggregation weights in their convolutional kernels for spatially neighbouring input features. The specific challenges or limitations posed by gradient exploding or vanishing and fixed aggregation weights in CNNs should be elaborated for better clarity.

To address these challenges, we propose a deep learning model called ReLATEC for diagnosing respiratory diseases from breath sounds collected with mobile devices. ReLATEC combines CNNs, RNNs and Transformers to analyse breath sounds and leverages local attention and recurrent structure on the Transformer model. This approach has shown more robust results compared with conventional methods.

The combination of these methods in ReLATEC leads to significant improvements in performance, memory usage, training time and overfitting compared to CRNN, Recurrent Transformer and vanilla Transformer models. ReLATEC outperforms both the vanilla Transformer and conventional CRNN approaches.

Our contribution lies in the design of ReLATEC, which enables precise diagnosis of respiratory diseases using breath sounds captured from mobile devices. By leveraging the strengths and addressing the weaknesses of CNNs, RNNs and Transformers, we enhance the accuracy of respiratory disease diagnosis.

4.2 Searching Health Information Using Breathing Sounds

4.2.1 Introduction

IoT applications for breath sounds have great potential to provide fast, low-cost, contactless and non-invasive monitoring of respiratory health problems. Respiratory sounds, or breath sounds, contain important indicators of respiratory illness (B. Lei et al., 2014; Marshall & Boussakta, 2007). Breath sounds can be non-invasively collected by handheld microphones (Bokov et al., 2016; I. Song, 2015). Previous studies using automatic breath sound diagnosis focused on detecting abnormal breath sounds including wheezes (Bahoura, 2009; Bokov et al., 2016; Naves et al., 2016; Taplidou & Hadjileontiadis, 2007a, 2007b), crackle (İçer & Gengeç, 2014; Naves et al., 2016) and rhonchi (İçer & Gengeç, 2014). Other studies focused on classification of medical conditions (Charleston-Villalobos et al., 2011; B. Lei et al., 2014; Sánchez Morillo, Astorga Moreno, Fernández Granero, & León Jiménez, 2013; I. Song, 2015).

However, the current approaches lack interpretable breath pattern abnormality analysis for further understanding of respiratory conditions. This hinders effective searches for health-related information within HSNs, requiring manual interpretation and monitoring by trained medical professionals (Manogaran et al., 2018).

To address the challenges of interpreting breath sounds and facilitating targeted information retrieval, we aim to transform breath sounds into interpretable clinical measurements that represent specific physical abnormalities. These clinical measurements will be used as search keywords within the HSNs, improving search capabilities and enabling targeted information retrieval.

PhysioVec consists of three modules: (1) LRT for biosignal segmentation, which performs the segmentation of biosignals; (2) Multivariate Radial-Basis Logistic Interpreter (MLI) for the physiological description of patients, which provides clinical descriptions based on vital respiratory physiological measurements; and (3) SBERT (Reimers & Gurevych, 2019) for semantic search of online content, which enables the searching of online documents. Breath sound recordings are collected non-invasively using a common mobile phone or digital recorder. The advantages or benefits of this non-invasive approach should be briefly explained. From the segmented breathing

sounds three types of vital respiratory physiological measurements, namely breath rate, breath pitch and breath depth, are extracted. MLI estimates five types of respiratory abnormalities, including tachypnea, hyperpnea, hypopnea, wheeze, and dyspnea, from vital respiratory physiological measurements. The clinical descriptions obtained are then used for searching online documents, providing valuable insights or information. Finally, SBERT, a pre-trained NLP model, is applied to convert both online documents and clinical descriptions into vectors for enhanced semantic searching.

LRT achieved a segmentation IoU of 77.0%, which was 24.0% higher than the conventional Transformer's performance of 62.1%. Additionally, it achieved Pearson scores of 91.71%, 72.10%, and 88.15% in estimating three physiological measurements: breath rate, breath pitch and breath depth, respectively. The predicted breathing abnormalities (disease) had an average sensitivity of 62.9% and an average specificity of 94.1%. In online searching tasks, the top-one search results achieved precision rates of 100%, 59.8%, 92.2%, and 100% for common cold, influenza, pneumonia and bronchitis, respectively. The precision of the top one for all conditions is 0.903. PhysioVect successfully integrated IoT with HSNs allowing patients and doctors to directly search relevant health information using breathing sounds. It also provides interpretable physiological and pathological information about underlying respiratory health conditions. It not only predicts the disease but also provides rich information to patients for them to better understand their health conditions with more clarity.

The major contributions of our approach are: (1) We proposed a novel LRT combining local attention and recurrent Transformer to reduce overfitting and improve performance in the segmentation of breathing sounds; (2) interpret breath pattern abnormality by extracting physiological measurements; and (3) PhysioVect successfully integrated IoT with HSNs allowing patients and doctors to directly search relevant health information using breathing sounds.

4.2.2 Method

4.2.2.1 Overall Process

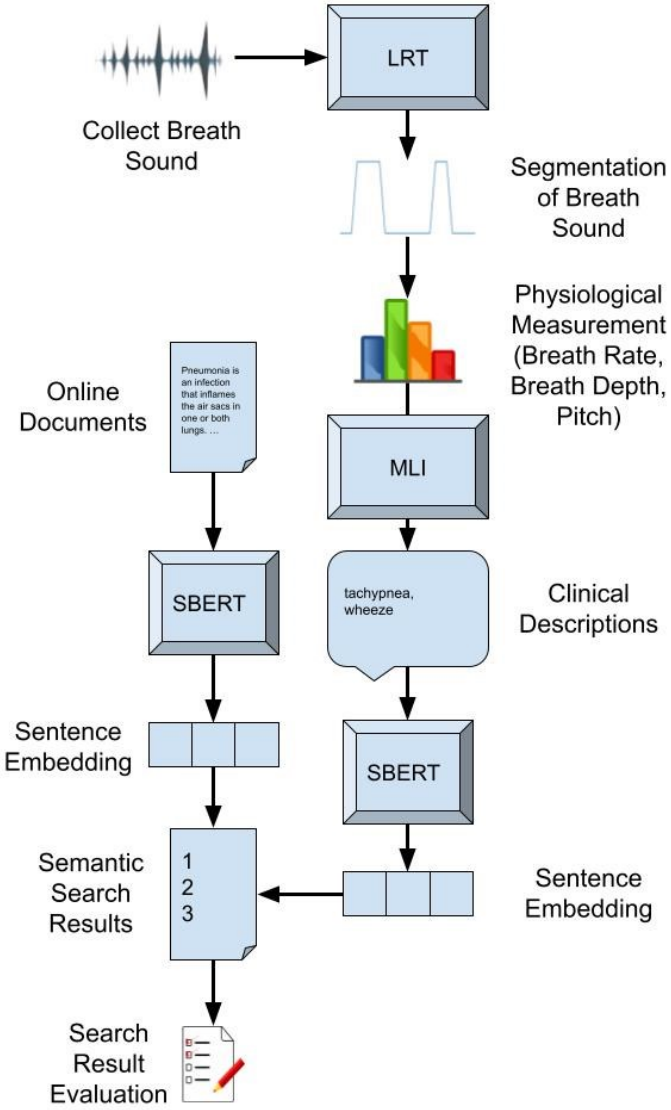

Figure 4.16 ***Overall Process of the proposed PhysioVec.***

Figure 4.16 illustrates the overall process of the proposed PhysioVec system. Breath sounds are collected non-invasively using common IoT or digital recorders. Next, the system generates a prediction of the segmentation annotation for the breath sound. From the segmented breath sound, physiological measurements are extracted. These measurements are then converted into clinical descriptions by MLI based on statistics and existing knowledge. Finally, the clinical descriptions, along with online documents, undergo vector conversion using SBERT. The resulting vectors enable semantic searching, facilitating the retrieval of relevant information.

4.2.2.2 Dataset

The breath sound recordings were collected from 1,144 patients from hospitals and clinics in Bangladesh (I. Song, 2015). The data were labelled with health conditions (cold, flu, bronchitis, and pneumonia), age and exhale cycles. Each exhale cycle was labelled with the beginning and ending time.

Data were divided into training sets; validation sets and test sets. The test set was selected with 20% random sampling. The training set and validation set were further selected with a proportion of 80%. After that, a 30-second-long segment was extracted from each case which contained at least three exhale cycles.

In this study, we aimed to collect information on four common respiratory diseases: the common cold, influenza, pneumonia, and bronchitis. To collect data, we searched for each disease name on Google and manually collected the results on the first page. We then used the physiological descriptions of each disease as search queries to locate relevant articles. In total, we collected 35 articles from various reputable sources on the internet.

The articles collected were thoroughly examined to extract information on the pathophysiology, symptoms, diagnosis, treatment, and prevention of each disease. We ensured that the articles collected were up-to-date and published within the past few years to ensure the validity and reliability of the information collected. We also made sure to include articles from various sources to ensure a comprehensive understanding of each disease.

4.2.2.3 Local Relation Transformer (LRT)

In our work, we developed and trained LRT, a customized Transformer model specifically designed for breath sound segmentation. In comparison to the conventional Transformer model, LRT incorporates both local attention and a Recurrent encoder block. These modifications were implemented to address overfitting issues and enhance the overall performance of the model.

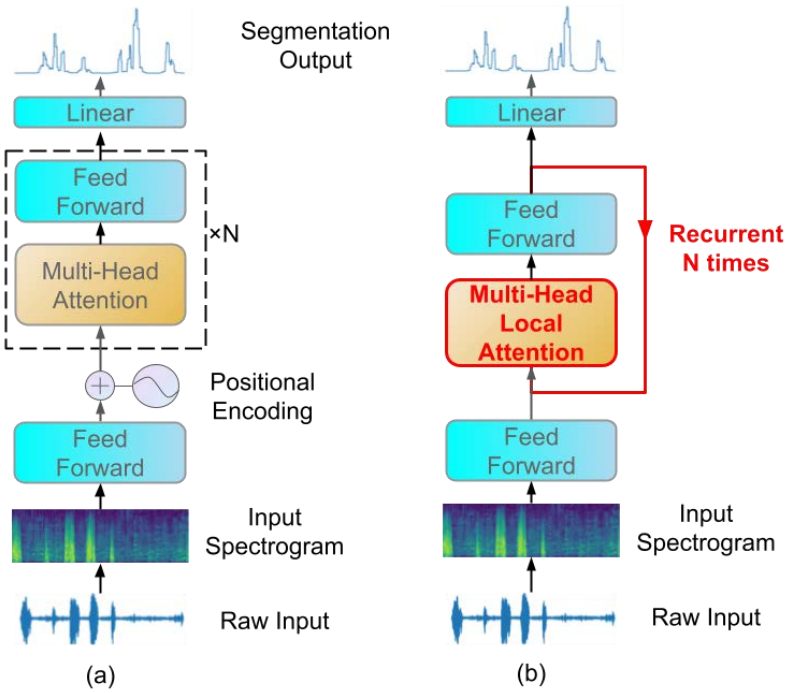

Figure 4.17 ***Comparison of vanilla Transformer (a) and ReLATEC (b).***

Figure 4.17 provides a comparison of the breath sound segmentation procedures using both LRT and Transformer models. The input feature maps for both models are spectrograms, which are derived from raw audio data. The entry block in the models serves as a simple feed-forward network. To enhance the robustness of the segmentation results, the hidden layer is configured with a dropout rate of 50% during training. The major difference between LRT and Transformer models is visually emphasized with bold red styling. Firstly, LRT utilizes a local relation approach to generate contextualized features, employing a local aggregation window of the input and a learnable local positional embedding, as opposed to relying on global relations. This allows for more localized feature generation. Additionally, LRT incorporates a recurrent encoder block to reduce memory usage, replacing the use of multiple individual encoder blocks. Finally, a linear layer is employed to convert the output features into the segmentation output.

The concept of local relation was initially introduced in LR-Net (Hu et al., 2019), an image deep-learning model that utilizes sliding window attention to extract contextual features based on the local relations of input features. The findings from LR-Net demonstrated that the incorporation of local relation introduces an information bottleneck to the feature learning process, which has proven to be crucial in computer vision tasks and leads to improved performance. Building upon this, we hypothesize that this information bottleneck effect of local relation can also enhance performance in the analysis of breath sound. Furthermore, the adoption of local attention reduces spatial complexity, allowing the model to effectively process longer sequences. In contrast, the Vanilla Transformer model calculates attention energy between all input nodes, resulting in a many-to-many relation. Given an input size of n, the Vanilla Transformer model exhibits a spatial complexity of $O(n^2)$ when

storing attention energy. However, with local attention, the attention energy is only calculated within a fixed-sized window, resulting in a complexity of $O(n)$. Multi-head local attention is computed using the method described in equations (4.10-4.12).

$$Attention(Q,K,V) = softmax\left(\frac{Q(K+P)^T}{\sqrt{d_k}}\right)V, \tag{4.10}$$

$$MHA(Q,K,V) = concat(head_0, \dots head_h), \tag{4.11}$$

$$head_i = Attention(QW_i^q, KW_i^k, VW_i^v), \tag{4.12}$$

where d_k denotes the number of embedding dimensions, W denotes learnable weights, P denotes learnable positional embedding, and i denotes the head number.

To address the issue of overfitting and enhance performance, we introduced a recurrent encoding block in our approach. Previous studies, such as UT and ALBERT, have provided evidence of the superiority of this strategy over vanilla transformers (Dehghani et al., 2018; Lan et al., 2019). In our LRT approach, rather than utilizing multiple independent layers, we iteratively apply a single encoder block six times, as shown in equations (4.13) and (4.14).

$$X_i = LRT(X_{i-1}), \tag{4.13}$$

$$LRT(X_i) = FFN(X_i + MHA(X_i, X_i, X_i)), \tag{4.14}$$

where X_i is output from iteration i; W and b denote learnable parameters; *FFN* is a feed-forward network with a hidden layer; *MHA* is multi-head local self-attention. The multi-head attention is used as self-attention, where query, key, and value are all input.

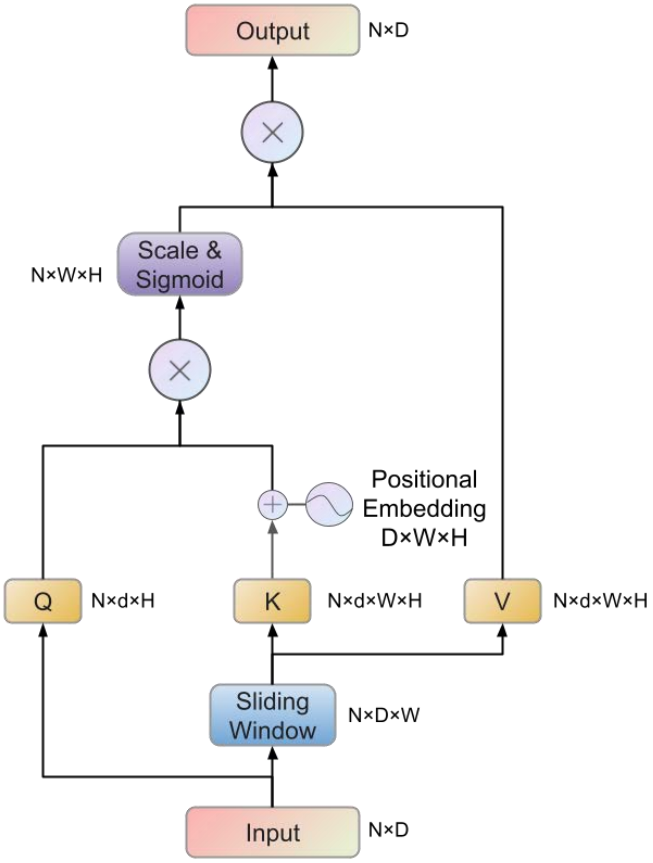

Figure 4.18 ***Local Attention Implementation in Our Approach.***

Figure 4.18 illustrated the implementation of multi-head local attention in our approach. The input size is $N \times D$, where N represents the length of the input sequence, and D denotes the dimension of the input feature. Sliding windows of size W are generated from the input, alongside the utilization of the dimension N for key and value associations. Through linear layers without bias, query, key, and value are obtained. Query, key, and value are then split into H heads. The energy is calculated as the dot product of the query and key, incorporating dimension d. The Softmax of scaled energy is then calculated, to build attention with a local aggression window for each head. Finally, the values for each head are computed as the dot product with dimension d, and then concatenated to form the output.

The loss function to train the model is frame-wise cross-entropy. Since some samples of the collected breath sounds are not fully annotated, the loss after the last annotated exhale cycles is set to zero. We trained the LRT and Transformer using a training configuration of 50 epochs and batch sizes of 64. The models were trained using the AdamW optimizer with an initial learning rate of 0.0001 and a weight decay of 0.0001. The learning rate was reduced by 20% after each epoch. Additionally, early stopping was implemented with a patience of 5 epochs to prevent overfitting.

4.2.2.4 Respiratory Phycological Measurements

Respiratory physiological measurements are extracted from the longest continuous exhale cycles of each breath sound recording. To find the continuous exhale cycles, we detect the outliers of inspiratory cycles with Tukey's Fence algorithm as (4.15).

$$T = Q_3 + 1.5(Q_3 - Q_1), \tag{4.15}$$

where Q_3 and Q_1 denote the upper quarter and lower quarter of the sample. Inspiratory cycle outliers are longer than T. A group of continuous exhale sounds does not contain an outlier inspiratory cycle. Then the longest continuous exhalation cycles are selected.

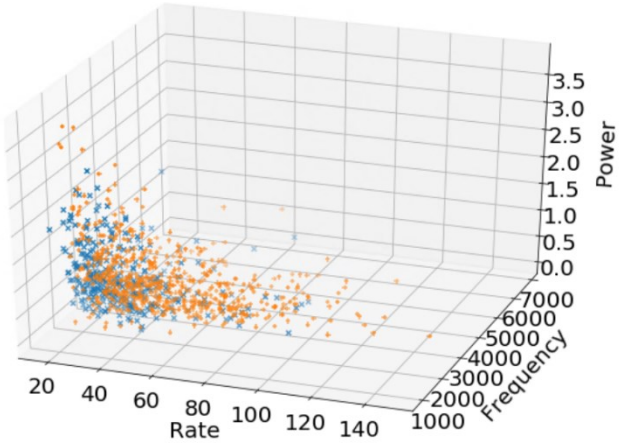

Figure 4.19 ***Plot of physiological measurements of breath sounds: breath rate, frequency (pitch), and power (depth). Healthy sounds are in blue and abnormal sounds are in orange.***

After that, we extracted four respiratory measurements from the segmentation of breath sounds: breath rate, breath depth and spectral centroids. Figure 4.19 plots the breath rate, frequency, and power of healthy (blue) and diseased (orange) breath sounds. Compared with healthy breath sounds, diseased breath sounds have lower power, higher pitch and higher breath rate.

Breath rate was calculated as (4.16):

$$Breath\ rate = \frac{S_n - S_1}{n-1}, \tag{4.16}$$

where S denotes the start of an exhale sound in the recording, i denotes the exhale sound number, and n denotes the total number of exhale sounds.

Mean breath depth is measured with the power of the audio recording, which was calculated as (4.17-4.19).

$$p_t = \sqrt{X_t^{\,2}}, \tag{4.17}$$

$$p_i = \sum_t^{E_i+S_i} P_t, \tag{4.18}$$

$$P = \frac{\sum_i^n p_i}{n}, \tag{4.19}$$

where *X* denotes the spectrogram of the sound, *S* denotes the start of the exhale sound in the recording, *E* denotes the end of the exhaling sound in the recording, *i* denotes the exhale sound number, and *n* denotes the total number of exhale sounds.

The pitch of a breath sound can be determined by analysing the mean exhale spectral centroid. Wheezing is characterised by a high-pitched breath sound, which often indicates a narrowing of the windpipe. Spectral centroid is calculated as (4.20).

$$C_t = \frac{\sum_i^n f_i X_{ti}}{\sum_i^n X_{ti}}, \tag{4.20}$$

where *X* denotes the spectrogram of the sound, *i* denotes the spectrogram number, $.X_{ti}$ denotes the energy in *i* frequency and *t* time of the spectrogram, f_i denotes the frequency of channel *i* of the spectrogram.

The average pitch of breath sound is calculated as (4.21) and (4.22)

$$c_i = \sum_t^{E_i+S_i} C_t, \tag{4.21}$$

$$C = \frac{\sum_i^n c_i}{n}. \tag{4.22}$$

where *S* denotes the start of the exhaling sound in the recording, *E* denotes the end of the exhaling sound in the recording, *i* denotes the exhaling sound number, and *n* denotes the total number of exhaling sounds.

Table 4.10 ***Normal Range for each Respiratory Physiology Measurement in Different Age Groups.***

Age Group	**Breath Rate (Breathing Per Minutes)**	**Pitch (Hz)**	**Breath Depth (db)**
<1	24-60	1160-2359	0.218-0.788
1-2	24-40	1160-2359	0.349-1.466
3-5	22-34	1160-2359	0.419-1.932
6	18-30	1160-2359	0.427-2.238

The measurements are converted into abnormal breathing patterns using a mapping function, a combination of Gaussian radial-basis, and logistic functions. Clinical normal distribution parameters are calculated from the healthy subjects in the training set. Clinical normal distribution parameters are

used to initialise the functions. Table 4.10 shows the normal range of each breath rhythm respiratory physiology measurement. Measurement within the normal range is considered normal. We use 1.65 Sigma to determine the normal range of breath sound, which is the decision boundary of abnormal values for balanced sensitivity and specificity.

Table 4.11 ***Definitions of Breath Rhythm Clinical Descriptions based on Respiratory physiology measurement.***

Medical Term of Breathing Abnormality	**Commonly used Synonym**	**Rate**	**Depth**	**Pitch**
tachypnea	rapid breath	High		
hyperpnea	deep breathing		High	
hypopnea	shallow breathing		Low	
wheeze	wheeze			High
dyspnea	Shortness of Breath	High	Low	

Abnormal breath patterns can be determined with a logistic function according to the Sigma of physiological measurements. The logistic function uses the definition of abnormal breathing patterns based on previous studies (Table 4.11). From the breath sounds, we extracted seven abnormal breathing patterns. The commonly used synonyms of those abnormal breathing patterns are used as keywords for searching since the synonyms are more common on the internet.

Since both shallow breath (hypopnea) and tachypnea suggest loss of lung volume, we use shortness of breath to represent when a patient has both tachypnea and hypopnea.

4.2.2.5 Evaluation Metrics

We evaluated the performance of our proposed model using Intersection over Union (IoU) calculated as (5.6).

$$IoU = \frac{TP}{TP+FP+FN}, \tag{5.6}$$

where *TP* is the number of correctly detected exhale frames, *FP* is the number of detected exhale frames that are not correct, and *FN* is the number of exhaling frames that are not detected.

We evaluated the performance of our proposed model to generate the physiological measurement using percentage Root Mean Square Error (RMSE), and Pearson product-moment correlation coefficients. RMSE and percentage MSE were calculated as (5.7) and (5.8).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(y_i - p_i)^2}, \tag{5.7}$$

$$Percentage\ RMSE = \frac{RMSE}{Mean(y)}, \tag{5.8}$$

where *i* denotes the number of samples, *y* denotes ground truth and *p* denotes the predicted value.

Pearson correlation coefficients are calculated as (5.9).

$$\rho_{x,y} = \frac{c(x,y)}{\sigma_x \sigma_y}, \tag{5.9}$$

where *c* is the covariance; *x* and *y* are two sets to be compared; σ is the standard deviation.

We evaluate the search results with a utility function that detects whether the searched articles mention the disease name of the breath sound. We should notice that the search queries are only descriptions of breathing patterns (Table 5.14). Thus, this utility function evaluates the effectiveness of the clinical descriptions since we do not train any model to detect disease names directly.

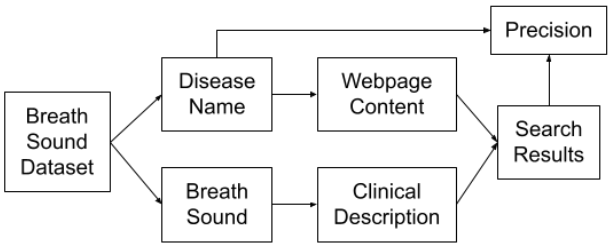

Figure 4.20 ***Process of evaluating the search results.***

Figure 4.20 shows the process of evaluating the search results. In the dataset, each biosignal is labelled with disease names. We also collect relevant articles for each disease. The clinical description is used to search among the collected articles. Articles ranked in the top 1, top 3, top 5 and top 10 will be retrieved as search results. After that, the utility function is used to evaluate the retrieved article.

We evaluate our search results based on accuracy and recall rates. Recall rate measures the percentage of correct articles ranked *n* retrieved from all correct articles. Given that we had more than five articles for each disease, we only calculated recall rates for the first 10 results.

4.2.3 Results

4.2.3.1 Breath Sound Segmentation

Table 4.12 ***Percentage RMSE of respiratory clinical physiological measurements prediction.***

Performance Measurements	**Transformer**	**LRT**
IoU (%)	0.621	0.770
Model Size (Megabyte)	1.950	0.124
Feed-Forward Memory Usage (Megabyte)	331.6	132.1

Table 4.12 presents the percentage RMSE of respiratory clinical physiological measurements prediction in the melt-down experiment. Even though LRT has lower feed-forward memory usage

(132.6M) than a normal vanilla Transformer (331.6M), LRT has overall better segmentation performance compared to the vanilla Transformer.

4.2.3.2 Respiratory Physiology Measurement Prediction Result

Table 4.13 shows the percentage RMSE of predicting respiratory physiology measurement with vanilla Transformer, and our proposed LRT. Table 4.14 shows Pearson product-moment correlation coefficients.

Table 4.13 ***RMSE of respiratory clinical physiological measurements.***

Respiratory Measurements	Transformer	LRT
breath rate (in seconds)	0.259	0.178
pitch	0.226	0.419
breath depth	0.376	0.402

Table 4.14 ***Pearson correlation coefficients of Respiratory Clinical Physiological Measurements.***

Respiratory Measurements	Transformer	LRT
breath rate (in seconds)	84.17%	91.71%
pitch	78.09%	72.10%
breath depth	85.01%	88.15%

Despite the pitch prediction being lower than the vanilla Transformer, the overall performance for predicting the breath rate and breath depth of the proposed LRT is outstanding. The reason for pitch having a lower correlation from the ground truth could be that the continuous exhale cycles for calculating pitch were different in predicted cycles and ground truth.

4.2.3.3 Clinical Description Prediction

Table 4.15 ***Performance of predicting clinical description.***

Name of abnormality	Accuracy	Sensitivity	Specificity
tachypnea	89.80%	88.10%	90.50%
hyperpnea	92.80%	33.30%	100.00%
hypopnea	88.30%	90.90%	87.90%
wheeze	74.50%	100.00%	63.50%
dyspnea	89.30%	90.00%	89.20%
average	86.94%	80.46%	86.22%

Table 4.15 presents the accuracy, sensitivity and specificity for predicting clinical descriptions based on predicted respiratory physiology measurements. In our study, we generated predicted abnormal respiratory patterns using the predicted respiratory physiology measurements. These predicted

patterns were then compared with the abnormal respiratory patterns obtained from the ground truth data. Our model achieved an average sensitivity of 80.5% and an average specificity of 86.2% for predicting seven different abnormalities.

4.2.3.4 Semantic Searching Result

Table 4.16 displays the precision of PhysioVec in searching for correct documents for specific breath sounds. The search results achieved top-one precision percentages of 100.00%, 59.80%, 92.20%, and 100.00% for breath sounds associated with the common cold, influenza, pneumonia and bronchitis, respectively. The overall top-one precision for all breath sounds is 90.30%. Notably, the results for pneumonia and the common cold demonstrate exceptional outcomes, with 100% precision for the first search results. These findings suggest that the semantic searching capability of PhysioVec provides potentially useful search results to users, despite using only a few disease descriptions as search queries.

Table 4.16 ***Precision of predicting diseases.***

Name of Disease	Top 1 Precision	Top 3 Precision	Top 5 Precision	Top 10 Precision
all	90.30%	66.50%	64.30%	60.40%
pneumonia	100.00%	67.60%	66.70%	68.90%
bronchitis	59.80%	43.10%	36.80%	40.90%
influenza	92.20%	67.30%	63.10%	39.20%
cold	100.00%	76.20%	84.80%	82.20%

Table 4.17 ***Recall rate of predicting diseases.***

Name of Disease	Recall Rate
cold	35.70%
influenza	30.20%
pneumonia	38.30%
bronchitis	40.90%

Table 4.17 shows the recall rate of documents of each disease. The search results achieved above 30% recall rate in searching documents for common cold, influenza, pneumonia and bronchitis, respectively. In a real scenario, a 30% recall rate means 30% of potentially useful online documents can be searched by clinical description. With the huge amount of data on the internet and HSNs, users can obtain sufficient information for their health conditions.

4.2.4 Conclusion

This section introduces PhysioVec, a multistage deep-learning framework for indexing breath sounds through searching HSNs directly using biosignals and solves research question $RQ_{2.2}$ and hypothesis

$H_{2.2}$. PhysioVec addresses the challenge of enabling the semantic search of text-based content from HSNs by converting biosignal data into clinically interpretable descriptions. These descriptions serve as keywords for searching relevant health information within HSNs. The main objective of this research is to develop methods that can effectively convert breath sounds into interpretable clinical measurements, describing abnormalities in physical measurements and providing keywords for semantic search within HSNs. PhysioVec employs the LRT to segment breath sounds, extract patterns and process biosignals to derive specific physiological measurements. By statistically analysing these measurements, PhysioVec generates human-readable keywords for semantic searching, improving search efficiency within HSNs. The contribution of this research is the development of PhysioVec, which enables semantic search of text-based content in HSNs using clinically interpretable descriptions derived from breath sounds. Furthermore, PhysioVec is extendable and customisable, facilitating automatic early diagnosis, comprehensive respiratory health tracking and visualisation of respiratory physiology measurements. Future research may explore incorporating additional respiratory physiology measurements, such as ECG data, patient description text and body temperature readings, to further enhance the effectiveness and usability of respiratory Health IoT applications.

5 Integrated HSN for Gastrointestinal Diseases

5.1 Bowel Sound Analysis with Bowel Sounds Sensor in Term Newborns

5.1.1 Introduction

Gastrointestinal problems such as diarrhoea, feeding intolerance, and Necrotising Enterocolitis (NEC) are major health threats to term newborn infants who are often cared for in the NICU. Currently, there are no known automated bowel activity monitoring systems available to detect potential health risks of term newborn infants resulting mortality rates of thousands per year in Australia alone.

In addition, disparities in accessing healthcare services between rural and urban areas contribute to limited treatment options. In developing countries, it has been reported that only 39% of children with diarrhoea receive the recommended treatment as per the guidelines set by the World Health Organization (WHO) (UNICEF & Diarrhoea, 2012). This lack of access to healthcare further exacerbates the issue, leaving many individuals in rural areas undiagnosed and untreated for gastrointestinal issues (B. Lei et al., 2014; I. Song, 2015).

However, the lack of affordable methods for collecting biosignals poses a challenge in detecting bowel diseases in newborns within the NICU. Currently, there is no known method for continuous monitoring of gastrointestinal health in term newborn infants at the NICU. In addition, locating and detecting intestinal sounds from digital recordings remains a challenge due to the irregular patterns and occurrence of intestinal sounds (Ranta et al., 2004; Sazonov et al., 2010). Nonstationary characteristics of bowel sounds in duration and amplitude variations make recognising bowel sounds difficult, especially for physicians (Hadjileontiadis & Rekanos, 2003).

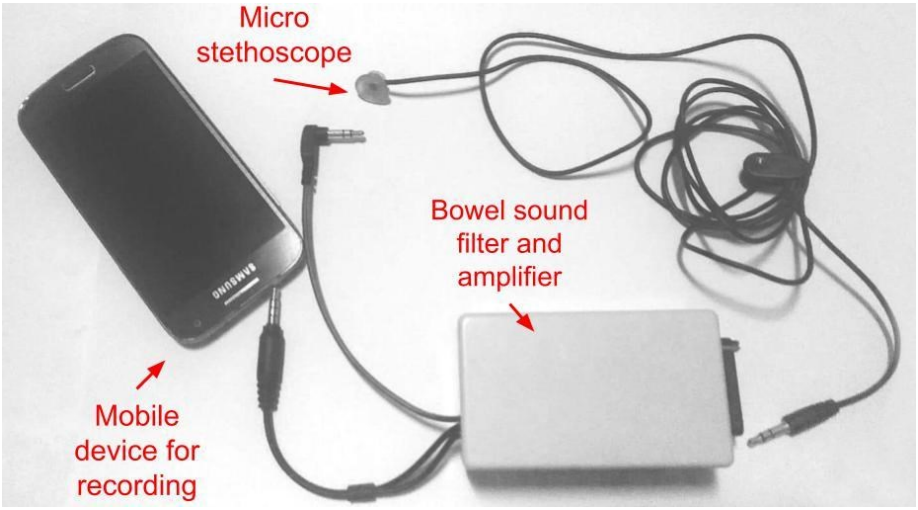

Figure 5.1 ***Miniature BoSS.***

To address the challenges, we aim to achieve real-time, non-invasive monitoring and diagnosis of gastrointestinal problems in newborn infants by automating the monitoring and detection of bowel activities. The goal is to develop cost-effective methods for collecting bowel sounds from newborn

and term infants within the NICU, extract clinical measurements from the bowel sounds, and perform subsequent exploratory data analysis.

Therefore, we developed an innovative miniature Bowel Sounds Sensor (BoSS), and automated detection and alert system of bowel sound (Figure 5.1). The 3D-printed micro stethoscope of BoSS allows continuous monitoring of bowel sounds of newborn infants with minimal discomfort. This non-invasive method reduces patient pain and does not require a trained physician (Y. Huang & Song, 2019; Y. Huang et al., 2017; I. Song, 2015).

We have collected a total of 240 bowel sounds from two NICU hospitals. We report on the data collected and physiological measurements that can be obtained from the bowel sound records done by BoSS. In this study, we annotated bowel sound events in newborn infants. Based on the annotation, we further perform bowel clinical analysis for term newborn infants. The result of this study is that we identified bowel sound physiological measurements that can be used for the diagnosis of gastrointestinal health.

In our study, we categorised bowel sounds into two distinct categories: ticks, which are characterised by short bursts, and grumbles, which are characterised by long bursts. To analyse these bowel sounds, we calculated physiological measurements for each recording, considering the labels of ticks and grumbles.

The physiological measurements were calculated using two different approaches: global and local. Global measurements were derived from the entire audio recording, while local measurements were obtained by considering the arithmetic median of sliding windows within each recording. The use of local measurements proved to be advantageous as it removed the silent portions of the recording, resulting in better quality measurements.

We conducted a thorough analysis by plotting and comparing the distribution of measurements across various factors such as gender, disease, age, and pre-feed/post-feed status. We aimed to demonstrate the feasibility of continuous monitoring of gastrointestinal health in term newborn infants.

In Section 8.5, Figure 5.17 presents a box plot illustrating the relationship between tick duration and conditions. The plot reveals that sick infants have lower tick durations compared with healthy infants. The average tick durations for healthy and sick infants are 0.177±0.0305 and 0.0998±0.0272, respectively. These differences are statistically significant at the 5% significance level. Similarly, in Section 8.5, Figure 5.18 displays a box plot of tick power versus conditions. The plot indicates that sick infants have lower tick power than healthy infants. The average tick power values for healthy and sick infants are 0.0991±0.0289 and 0.196±0.0294, respectively, and the differences are statistically significant at the 5% significance level. Furthermore, in Section 5.1.4, Figure 5.19 exhibits a box plot depicting grumble duration across different conditions. The plot demonstrates that sick infants have higher grumble durations compared with healthy infants. The average grumble durations for healthy

and sick infants are 0.0803±0.0161 and 0.131±0.0179, respectively, and the differences are statistically significant at the 5% significance level.

The results revealed clear differences in specific measurements between sick and healthy infants. Specifically, we observed significant differences (P=0.05) in tick rates, tick energy and tick durations. Sick infants exhibited lower tick (short burst bowel sound) activities and higher grumble (longer burst bowel sound) activities compared with their healthy counterparts.

Furthermore, the disparity between sick and healthy infants was more significant than the difference between pre-feed and post-feed conditions. For instance, the average values of tick durations were 0.177±0.0305 and 0.0998±0.0272 for healthy and sick infants, respectively. These differences were found to be statistically significant (P=0.05). Additionally, both pre-feed and post-feed tick duration and power were higher in the healthy group compared with the sick group.

Notably, there was a significant difference observed in grumble power between healthy and sick infants. The average values of grumble durations were 0.0991±0.0289 and 0.196±0.0294 for healthy and sick infants, respectively. Similarly, the average values of grumble power were 0.0803±0.0161 and 0.131±0.0179 for healthy and sick infants, respectively. These differences were found to be statistically significant with a significance level of 5%. These findings indicate that sick infants have longer and more powerful grumble activities compared with healthy infants.

The results of our study indicate that there are statistically significant differences (P=0.05) in several measurements between sick and healthy infants. Specifically, we observed significant differences in bowel sound rates, bowel sound energy and bowel sound durations. Sick infants exhibited lower levels of tick activities, which are characterised by short bursts of bowel sounds. On the other hand, sick infants demonstrated higher levels of grumble activities, which are characterised by longer bursts of bowel sounds. These findings suggest that there are distinct patterns in bowel sound activities between sick and healthy infants.

The main contributions of this study are as follows. First, this is the first time researchers have recorded bowel sounds of newborn infants during pre-feed and post-feed using miniature IoT sensors continuously. Second, we measure five clinical bowel activities and classify their types into short bursts (ticks) and long bursts (grumbles). Thirdly, we use an arithmetic median (median without duplicates) for filtering the physiological measurements of bowel sound activities for hospital environments and sensor misplacement and alignments. Fourthly, we find five physiological measurements that are significantly different between sick and healthy infants: tick rate, tick duration, tick power, grumble duration and grumble power.

5.1.2 Background

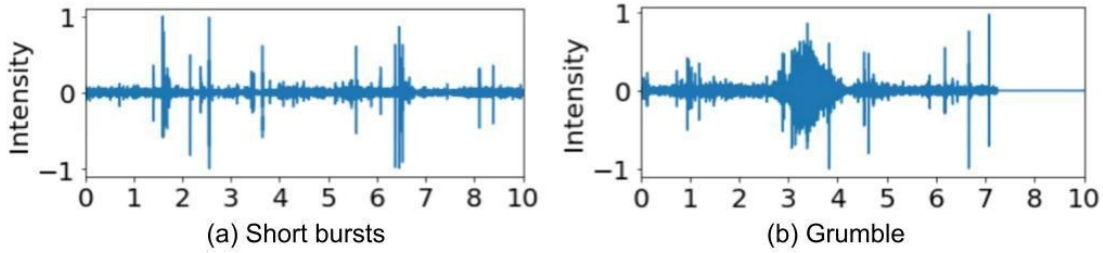

Figure 5.2 ***Examples of (a) short burst bowel sound and (b) long burst (grumble) bowel sound.***

Bowel sounds are usually high-pitched, so the diaphragm of a stethoscope should be used. Bowel sounds can be classified as normal, hyperactive or hypoactive. In this study, we identify two types of bowel sounds, short burst and long burst. Figure 5.2 shows examples of the two types of sounds. Short bursts have a higher pitch, shorter duration and sharp shape, which are also known as tickles, and clicks. On the other hand, grumble sounds have a lower pitch and longer duration, which are also known as clusters and grumble. In this study, we use the terms "tickle" and "grumble" to represent these two types of bowel sounds.

Previous studies have analysed bowel sounds by extracting jitter and shimmer for non-invasive bowel activity (K. S. Kim et al., 2011) monitoring and have compared similarities between bowel sounds (Y. Huang et al., 2017). There have also been studies focusing on bowel sound detection for long-term monitoring of gastrointestinal motility (Goto et al., 2015; Yin et al., 2018). Automated diagnosis based on biosignals has been applied to bowel diseases such as irritable bowel syndrome (Dimoulas et al., 2008) and abnormal bowel sounds (Y. Huang et al., 2017). Furthermore, bowel sound detection has been used to measure the frequency of bowel sounds (Sazonov et al., 2010; Yin et al., 2018). Mamun and McFarlane (2016) proposed an integrated bowel sound detection system using a hardware approach, which achieved a detection rate of 85%. This approach utilises a compact device for detection and offers faster processing due to its hardware-based implementation.

5.1.3 Data Collection

5.1.3.1 Recording

Table 5.1 ***Clinical Characteristics of the Study Patients.***

Age in days, range	**0-14**
Data source	
KK	191
Townsville	49
Gender	
Male	115
Female	125

Pre-feed or Post-feed	
Pre-feed	128
Post-feed	112
Health Condition	
Sick	155
Healthy	85
Detailed Bowel Health Conditions	
Surgical condition	14
Congenital anomaly	41
IUGR	18
LGA	20
TTS	0
APH	0
Cord Prolapse	0
Gastroschisis	0
Oesophageal atresia	2
Bowel obstruction	11
Bowel resection	0

The present study aimed to investigate the clinical characteristics of infants aged 0-14 days (Table 5.1). The data for this study was collected from two hospitals, namely KK Women's and Children's Hospital in Singapore and Townsville Medical Centre in Australia. The doctors and nurses of the hospitals were responsible for data collection. Over a period of five months, a total of 240 patients were included in the study. The gender distribution of the cohort was relatively balanced, with 115 males and 125 females. Bowel sound recordings were collected during pre-feed and post-feed in separate sessions for each baby. Most of the infants were sick (*n*=155) as opposed to healthy (*n*=85). Among the sick infants, the most common detailed bowel health conditions were congenital anomaly (*n*=41), bowel obstruction (*n*=11), and surgical condition (*n*=14), whereas Intrauterine Growth Restriction (IUGR) (*n*=18), Large for Gestational Age (LGA) (*n*=20), and oesophageal atresia (*n*=2) were less frequent. Notably, no cases of Twin Transfusion Syndrome (TTS), Antepartum Haemorrhage (APH), cord prolapse or gastroschisis were observed. The pre-feed and post-feed status of the infants was recorded, with 128 infants categorised as pre-feed and 112 as post-feed. These results provide valuable insight into the clinical characteristics of a relatively large cohort of newborns, which may inform clinical decision-making and guide future research efforts.

Table 5.2 ***Patient Statistics Sick/Healthy vs Pre/Post Feed.***

Dataset	Pre-feed	Post-feed	Total
sick	84	69	155
healthy	44	43	85
total	**128**	**112**	240

Table 5.2 presents the patient statistics according to the health status and feeding status of the cohort. The data reveal that among the sick infants, 84 were pre-feed and 69 were post-feed, while among the healthy infants, 44 were pre-feed and 43 were post-feed. Overall, most infants were sick and pre-feed (*n*=84), followed by sick and post-feed (*n*=69), healthy and pre-feed (*n*=44), and healthy and post-feed (*n*=43). The total number of pre-feed infants was 128, while the total number of post-feed infants was 112.

5.1.3.2 Data Labelling

A total of nine Masters in IT students specializing in this field from James Cook University were involved in annotating the bowel sounds. These students were trained on different types of bowel sounds, including both real and simulated samples, without knowledge of the patients' health conditions to avoid bias in labelling.

The annotation process involved two stages. In the first stage, a group of five students performed the initial labelling of the bowel sounds, followed by cross-verification and reclassification to ensure accuracy. In the second stage, the remaining four students independently revalidated the labels to further enhance the reliability of the annotations.

During the annotation process, the start and end time locations of bowel activities were manually marked, and each bowel sound was classified into two types: short bursts (ticks) and long bursts (grumbles). Short bursts were characterized by higher pitch, shorter duration, and sharp shape, while grumble sounds had a lower pitch and longer duration.

Table 5.3 ***Dataset after removed bowel sounds less than 10 ticks and without grumbles.***

Dataset	Pre-feed	Post-feed	All
sick	43	39	82
healthy	31	26	57
total	**74**	**65**	139

To analyse bowel activities related to ticks and grumbles, we filtered out recordings with fewer than 10 ticks or recordings without grumbles. The resulting dataset (Table 5.3) includes bowel sounds that meet the specified criteria.

5.1.4 Physiological Measurements

We selected five physiological measurements to measure bowel activities: tick rate, tick duration, tick power, grumble duration and grumble power. The tick rate is calculated for two types of segments: global and local. Global segment refers to the entire duration of a record. Local segment refers to a sliding window of a fixed size. In our study, the length of the sliding windows is 10 seconds.

We selected three features to measure bowel activities, which are tick rate, tick duration, tick power, grumble duration and grumble power. The tick rate is calculated as (5.1).

$$R_t = \frac{N_t}{T}, \tag{5.1}$$

where R_t is the tick rate, N_t is the number of ticks and T duration of the selected segment.

Tick duration is calculated as (5.2).

$$D_t = \frac{\sum E_i - S_i}{T}, \tag{5.2}$$

where D_t is tick duration, S_i and E_i are the start and end time of i-th tick, and T is the duration of the selected segment.

Grumble duration is calculated as (5.3).

$$. D_g = \frac{\sum E_i - S_i}{T}, \tag{5.3}$$

where D_g is grumble duration, S_i and E_i are the start and end time of i-th grumble, and T is the duration of the selected segment

Tick power is calculated as (5.4).

$$P_t = \frac{\sum P_i}{T}, \tag{5.4}$$

where P_t is tick power, P_i is the energy of i-th tick, and T is the duration of the selected segment.

Grumble power is calculated as (5.5).

$$. P_g = \frac{\sum P_i}{T}, \tag{5.5}$$

where P_g is grumble power, P_i is the energy of i-th tick and T is the duration of the selected segment.

In our study, we calculated physiological measurements using two different approaches: global and local measurements. In the global measurements, the duration of the selected segment was

determined as the length of the entire bowel sound record. On the other hand, in the local measurements, the duration of the selected segment was fixed at 10 seconds.

To obtain the final local measurements, we calculated the medians of the local measurements. The use of medians helped filter out outliers and noises that could arise from sensor misplacements and other environmental factors. Additionally, for the median calculation, we employed a method called arithmetic median, which involves calculating the median without including duplicate values.

By employing this approach, we aimed to obtain more robust and reliable measurements that accurately represent the physiological characteristics of bowel sounds.

Recording noise posed a challenge in capturing every bowel event during the recordings, while the non-stationary nature of bowel sounds made it difficult to record all events accurately. To address this, we employed the calculation of arithmetic median values for local clinical measures to overcome these challenges.

Calculating arithmetic median values for local clinical measures addresses the challenge of recording noise and the non-stationary nature of bowel sounds. It filters out extreme values caused by noise and focuses on relevant segments within the recordings. This approach provides more reliable measurements that accurately represent the physiological characteristics of bowel activities.

To calculate arithmetic median values for local clinical measures, we employed an automated process using sliding windows. For each subject, regions of interest (ROIs) within the bowel sounds were identified using sliding windows with a window size of 10 seconds and a stride of 0.1 seconds. The clinical measures within these ROIs were then used to calculate the arithmetic median as the local values.

To obtain the arithmetic median, any duplicate values from the sliding windows were removed. This step helped eliminate redundant or repeated measurements. Finally, the median was calculated based on the remaining values without duplication. The following figures plot the local values of the physiological measurements. The following figures plot the local values of the physiological measurements.

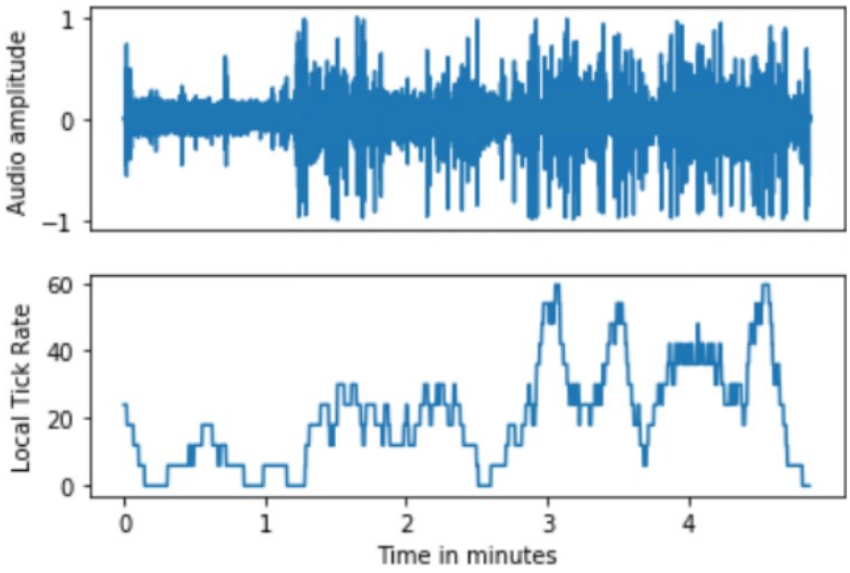

Figure 5.3 ***Local tick rate of a bowel sound recording.***

Figure 5.3 illustrates the local tick rate of an example recording. In this recording, the median tick rate is 18, the global tick rate is 20, and the arithmetic median is 30. However, the initial part of the audio contains insufficient bowel sound data, which cannot accurately represent the actual bowel activity. After approximately one and a half minutes, the arithmetic median approaches an average tick rate of around 30 per minute.

5.1.5 Results

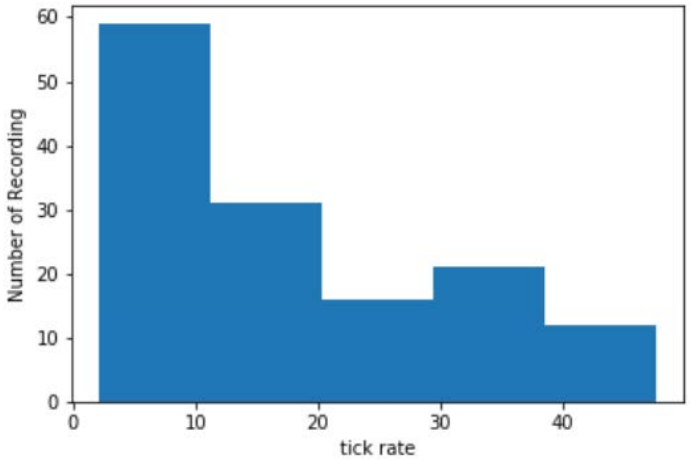

Figure 5.4 ***Global tick rate for all subjects.***

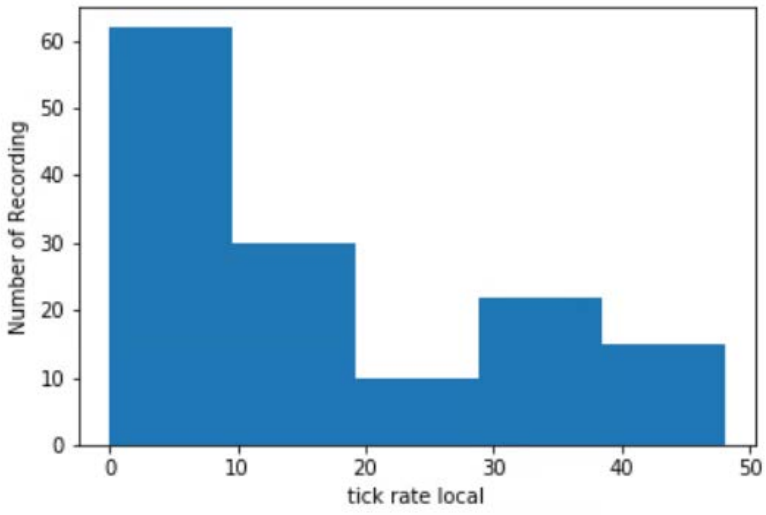

Figure 5.5 ***Local tick rate calculated with median.***

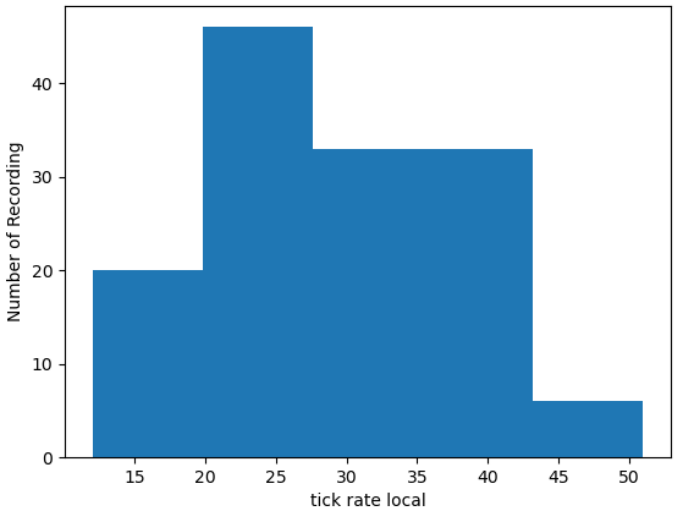

Figure 5.6 ***Median of local tick rate for all subjects after removing duplicated values.***

We now compare the global tick rate with the local tick rate. Figure 5.4 displays the global tick rate, which exhibits a high number of cases near zero. To address this, we remove the instances with zero values and retain the peak values observed around 30-40 ticks per minute. Figure 5.5 depicts the histogram of the local rate, which shows no significant difference compared with the global tick rate. The use of median values does not effectively eliminate the silence parts, as they typically account for more than half of our recordings. Figure 5.6 illustrates the local tick rate after removing duplicated values from sliding windows. Most cases demonstrate tick rates ranging from 15-40 ticks per minute, and the histogram exhibits a bell-shaped distribution pattern resembling a normal distribution.

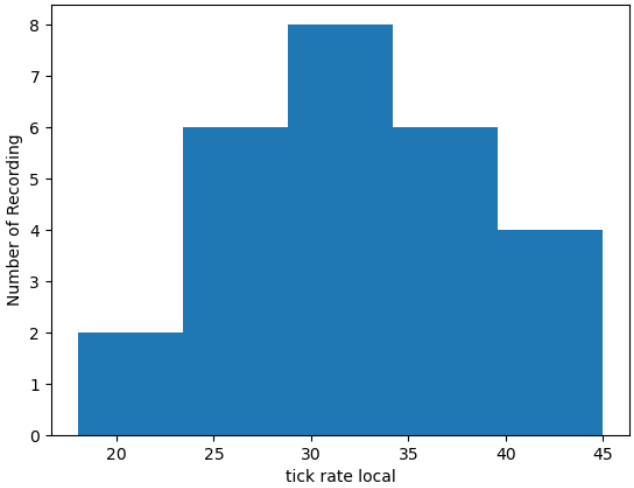

Figure 5.7 ***Local tick rate for healthy pre-feed.***

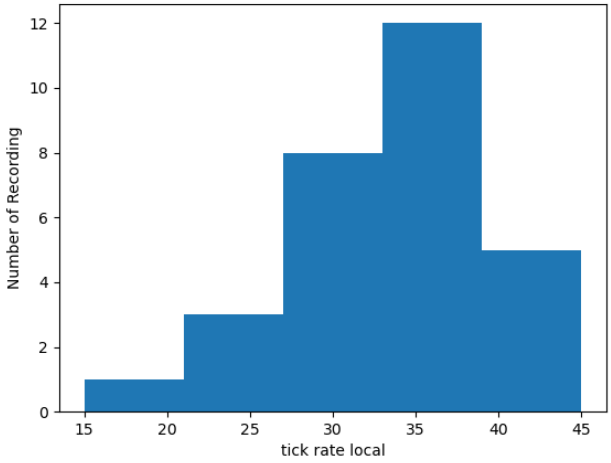

Figure 5.8 ***Local tick rate for healthy post-feed.***

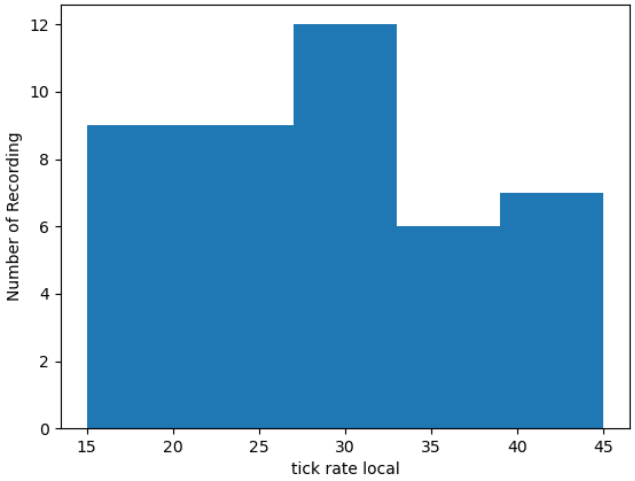

Figure 5.9 ***Local tick rate for diseased pre-feed.***

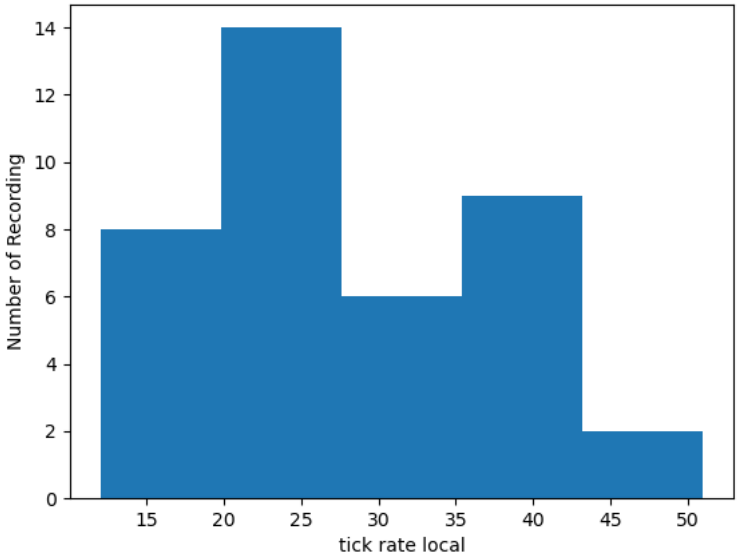

Figure 5.10 ***Post-feed sick local tick rate.***

Figures 5.7 and 5.8 illustrate the local tick rate for healthy subjects before and after feeding. It is observed that healthy individuals tend to have higher tick rates before feeding. On the other hand, Figures 5.9 and 5.10 display the local tick rate of sick bowel sounds before and after feeding. In the case of sick individuals, the tick rates vary significantly, with some recordings having too few or too many ticks. Despite having a larger number of recordings, the histogram for sick bowel sounds does not exhibit a bell-shaped distribution like the healthy group, especially in the pre-feed recordings. This suggests that bowel disorders can affect tick-related activity in different ways compared with pre-feed bowel sounds in healthy individuals.

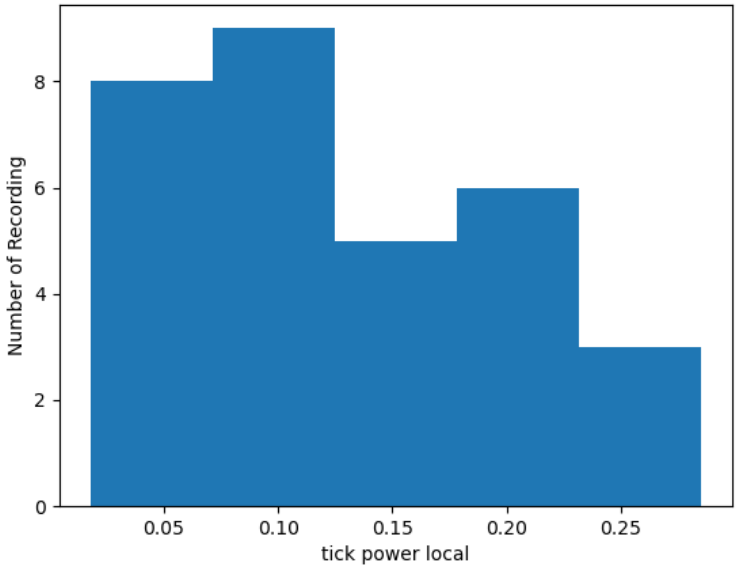

Figure 5.11 ***Pre-healthy tick power.***

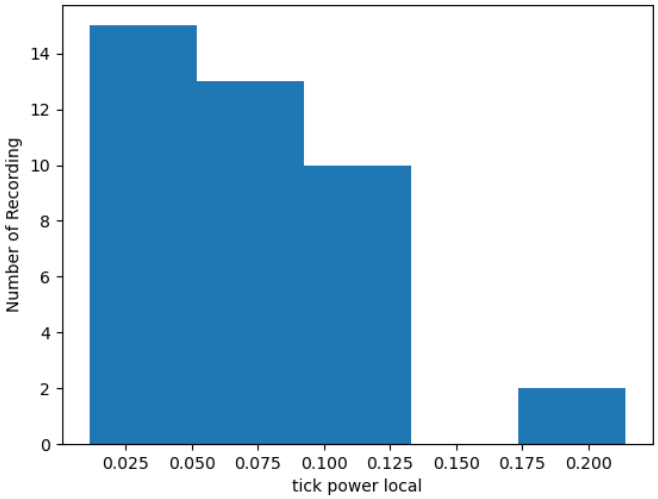

Figure 5.12 ***Pre-sick tick power.***

Power is another measurement of bowel activities. Compared with the rate which measures the frequency of bowel activity, power also indicates the intensity of bowel activity. Figures 5.11 and 5.12 show the histogram of the local power of healthy pre-feed bowel sound and sick pre-feed bowel sound. Sick bowel sounds have significantly less powerful tick sound or no tick sound.

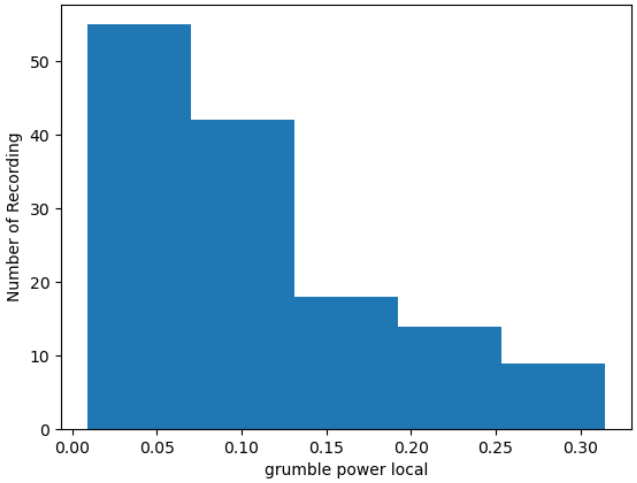

Figure 5.13 ***Grumble power for all.***

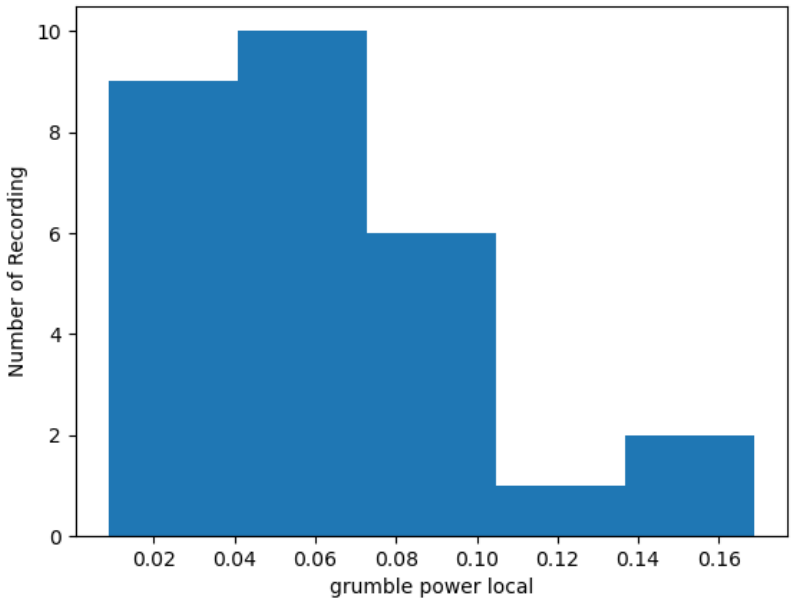

Figure 5.14 ***Grumble power for healthy pre-feed.***

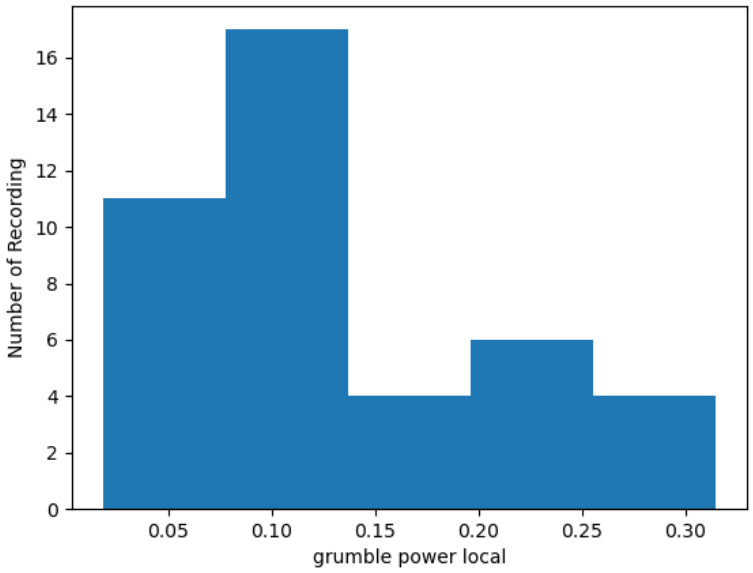

Figure 5.15 ***Grumble power for sick pre-feed.***

Figures 5.13, 5.14, and 5.15 show the grumble power for all recordings, healthy prefeed and sick prefeed. Since grumble is not usually seen in our dataset, some of the subjects have few grumbles, especially healthy subjects. By contrast, sick subjects have much more grumble. As a result, grumble could be a significant biomarker of sickness.

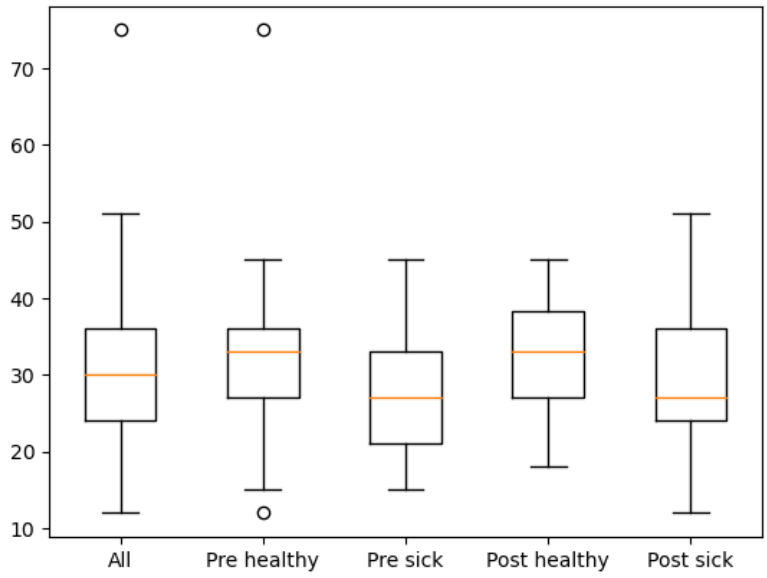

Figure 5.16 ***Box plot of local tick rate vs condition.***

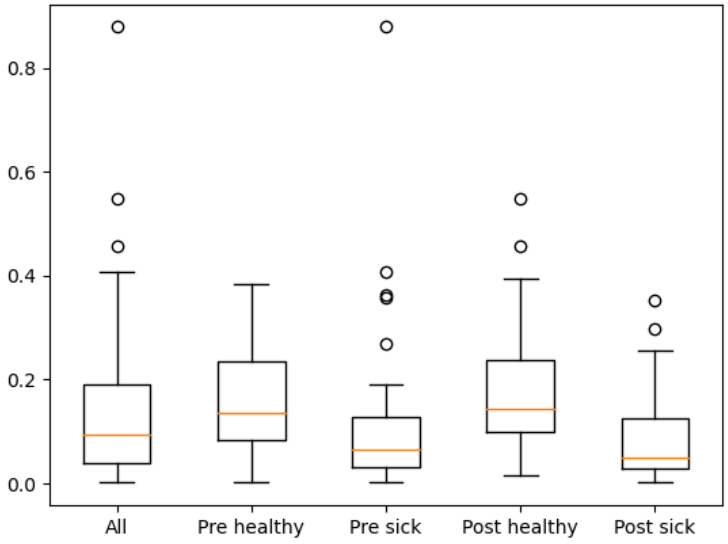

Figure 5.17 ***Box plot of tick duration vs conditions.***

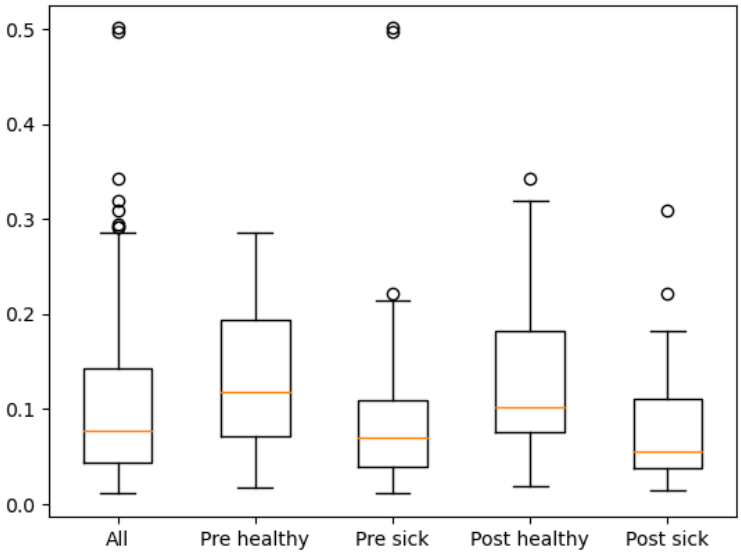

Figure 5.18 ***Box plot of tick power vs conditions.***

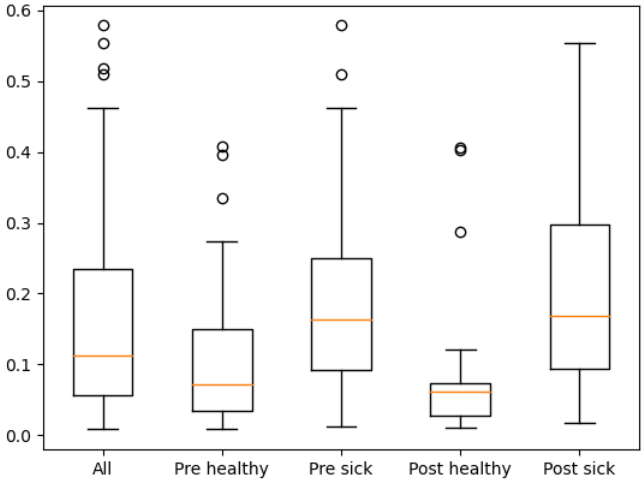

Figure 5.19 ***Box plot of grumble duration vs condition. Sick infants have higher grumble durations than healthy infants.***

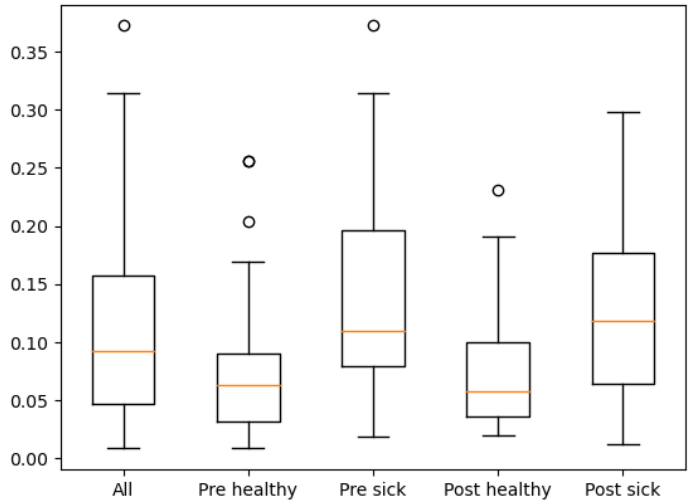

Figure 5.20 ***Box plot of grumble power.***

Table 5.4 ***Confidence intervals local measurements of pre and post-feeds of sick and healthy infants. The differences between sick and healthy are significant with P=0.05.***

	Healthy Pre-feed	**Healthy Post-feed**	**Sick Pre-feed**	**Sick Post-feed**	**Significant (*P*=0.05)**
Tick rate	32.8±3.83	32.4±2.81	28.0±1.87	28.8±2.94	Y
Tick duration	0.173±0.0371	0.182±0.0507	0.0998±0.0272	0.0884±0.0275	Y
Tick power	0.133±0.0260	0.142±0.0369	0.0860±0.0186	0.0789±0.0204	Y
Grumble duration	0.111±0.0403	0.0856±0.0416	0.196±0.0294	0.196±0.0418	Y
Grumble power	0.0823±0.0236	0.0780±0.0220	0.131±0.0179	0.129±0.0248	Y

Table 5.5 ***Confidence interval of merged local physiological measurements (pre and post-feed measurements are merged with weighted average) (P=0.05).***

	Healthy minimum	**Healthy maximum**	**Sick minimum**	**Sick maximum**	**Significant (*P*=0.05)**
Tick rate	30.2	35.0	26.2	29.9	Y
Tick duration	0.146	0.208	0.0725	0.130	Y
Tick power	0.115	0.159	0.0674	0.105	Y
Grumble duration	0.0702	0.128	0.167	0.226	Y

Grumble power	0.0642	0.0964	0.113	0.149	Y

Table 5.6 ***Another view of Table 8.4. Confidence interval of merged local physiological measurements (pre and post-feed measurements are merged with weighted average) (P=0.05).***

	Healthy	**Sick**
Tick rate	32.6±2.41	28.0±1.87
Tick duration	0.177±0.0305	0.0998±0.0272
Tick power	0.137±0.0219	0.0860±0.0186
Grumble duration	0.0991±0.0289	0.196±0.0294
Grumble power	0.0803±0.0161	0.131±0.0179

We then further analyse tick characteristics in different groups. Table 5.4 shows the confidence intervals of local physiological measurements in these groups. Sick infants have shorter tick durations compared with healthy infants. Figures 5.16 to 5.18 depict box plots of local tick rate, tick duration and tick power with the condition. They reveal that sick bowel sounds have a lower tick frequency, while post-feed bowel sounds have a slightly lower tick frequency. Furthermore, Figure 5.17 indicates that being sick is associated with both a lower tick frequency and weaker tick sounds. The mean of sick tick power is lower than the confidence interval (P=0.05) of healthy in both post-feed and pre-feed. The difference between sick and healthy is more significant than the difference between pre-feed and post-feed. The mean of sick grumble power is higher than the confidence interval (P=0.05) of healthy in both post-feed and pre-feed.

Table 5.5 and Table 5.6 show the confidence intervals of the merged local physiological measurements after merging the pre-feed and post-feed measurements using a weighted average. The average values of tick durations are 0.177 ± 0.0305 seconds and 0.0998 ± 0.0272 seconds for healthy and sick infants, respectively. Tick duration and power for healthy infants are higher than those in the sick group in both pre-feed and post-feed conditions. Notably, the minimum and maximum values for the healthy and sick groups are shown in bold font to emphasize that they do not overlap. Figure 5.19 displays a box plot of grumble duration vs condition. The average values of grumble durations are 0.0991 ± 0.0289 seconds and 0.196 seconds ± 0.0294 seconds for healthy and sick infants, respectively. There is a significant difference in grumble power between healthy and sick infants. Figure 5.20 shows a box plot of grumble power vs condition. The average values of grumble power are 0.0803 ± 0.0161 dBm s^{-1} and 0.131 ± 0.0179 dBm s^{-1} for healthy and sick infants, respectively. The results indicate that sick infants have longer grumble durations and higher grumble power compared to healthy infants (P=0.05).

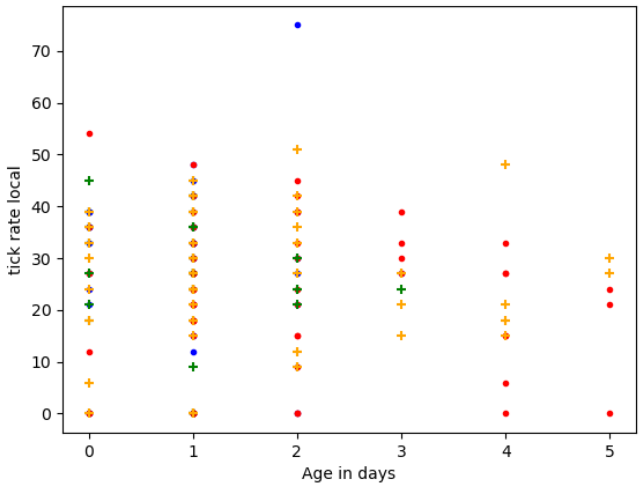

Figure 5.21 ***Local tick rate in different ages.***

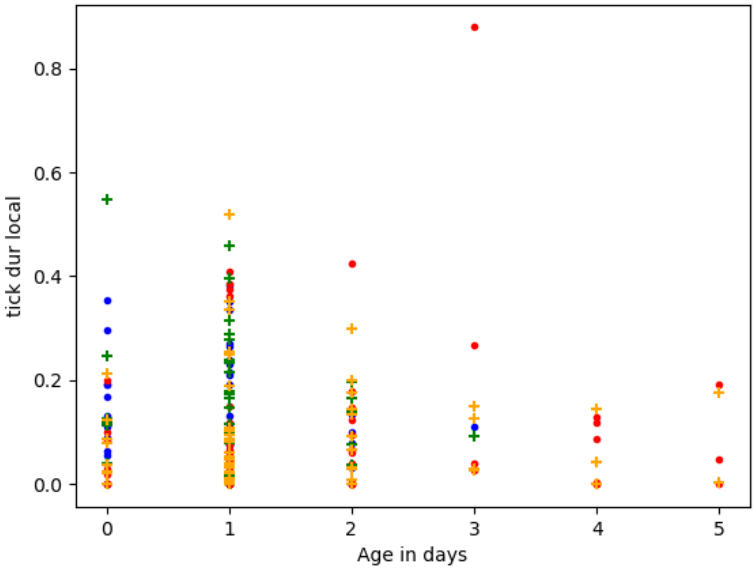

Figure 5.22 ***Local tick duration vs age.***

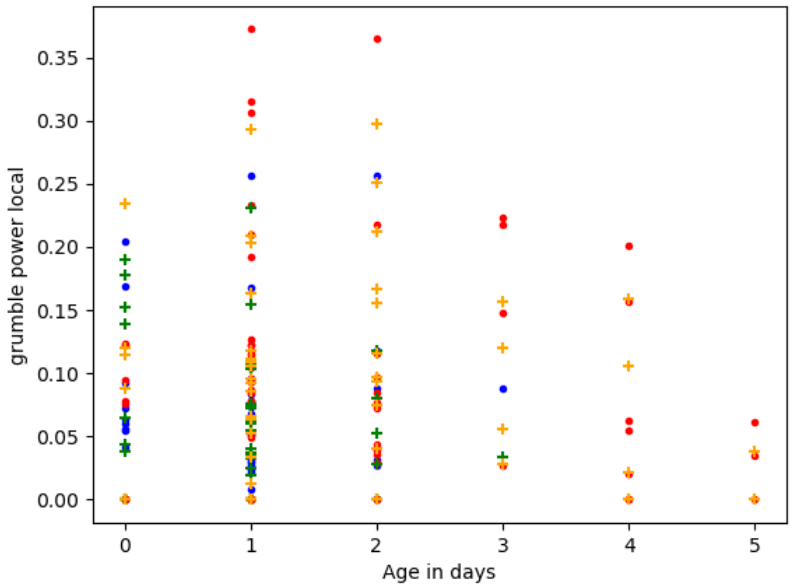

Figure 5.23 ***Local tick power vs age.***

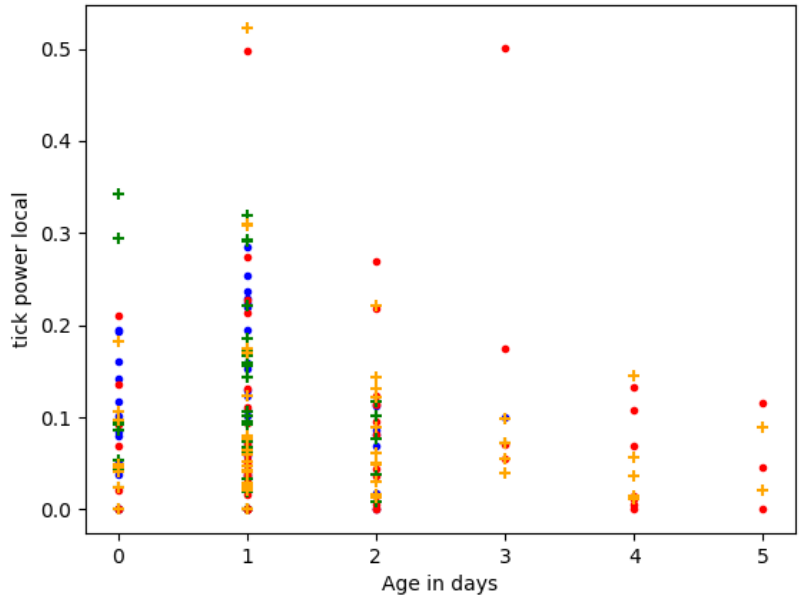

Figure 5.24 ***Local grumble duration vs age.***

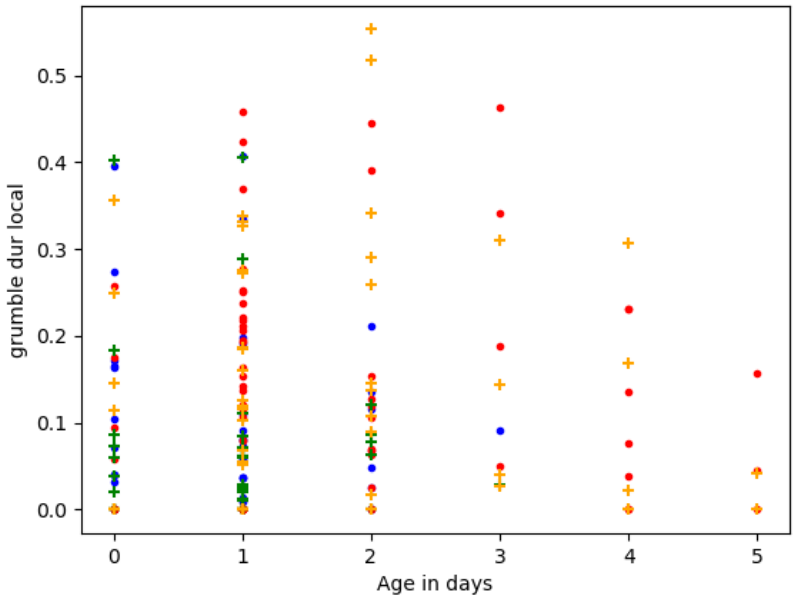

Figure 5.25 ***Local grumble power vs age.***

Figures 5.21 to 5.25 show scatter plots depicting the relationship between each measurement and age. Blue marks represent healthy pre-feed subjects, green marks represent healthy post-feed subjects, red marks represent healthy post-feed subjects, and yellow marks represent sick post-feed subjects. Most healthy subjects are around one day old and exhibit a tick rate of 30-50 ticks per minute. However, there is significant overlap among the groups, and the tick rates do not vary significantly with age.

In contrast to the local tick rate, the duration of local ticks shows a more significant difference between the healthy and sick groups. Therefore, local tick duration provides a better measurement for assessing bowel activity, which is related to the ticking sound. Furthermore, the duration exhibits a decreasing trend as age increases.

Power is another measurement of bowel activity that considers both the amount and intensity of the tick sounds. Compared with duration, the power values for pre-feed healthy subjects tend to be similar. Most recordings from the post-feed healthy group show a significant decrease in grumble duration, while most recordings from the sick group have a significantly longer grumble duration. However, grumble duration does not exhibit a significant trend with age. Power exhibits a decreasing trend across all groups with age, with the healthy pre-feed and post-feed groups being closer in values.

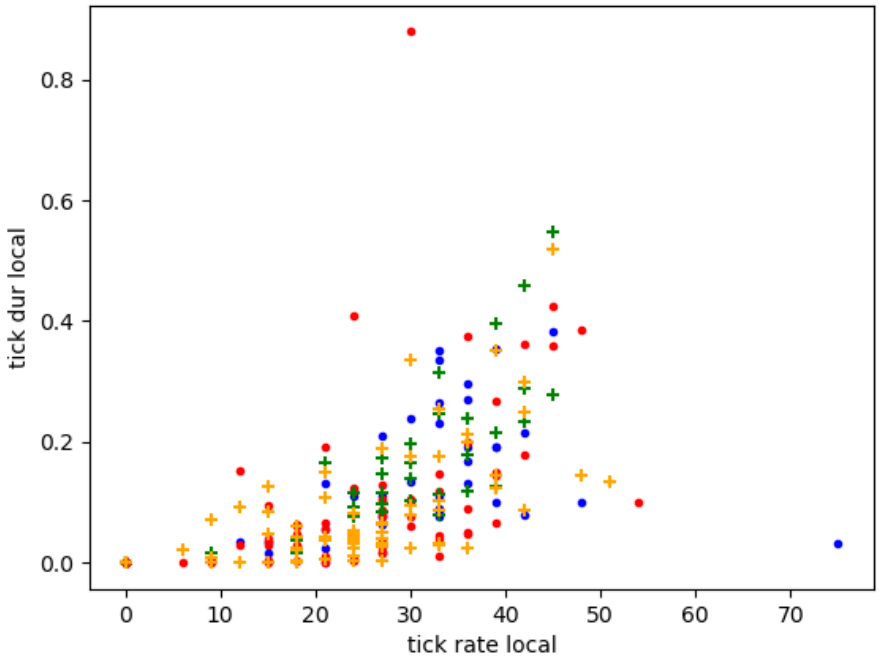

Figure 5.26 ***is the scatter plot between tick rate and tick duration.***

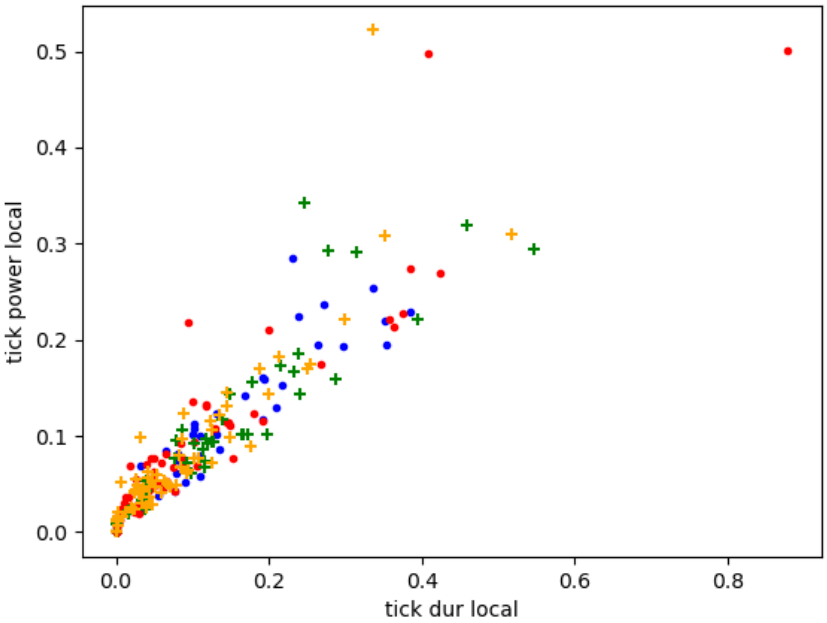

Figure 5.27 ***Scatter of tick power and duration.***

Figures 5.26 and 5.27 illustrate the relationship between tick rate, tick power and tick duration. Tick duration is highly correlated with tick rate, as most disease group subjects exhibit longer durations for the same tick rate. Since calculating tick duration does not require splitting ticks that are too close together, it can serve as a good replacement for tick rate as a measurement of bowel activity.

Additionally, there is a strong correlation between power and duration, indicating that ticks have stable power levels. Therefore, power can be used as an alternative measurement for bowel activity, as it is less sensitive to the precise duration of ticks.

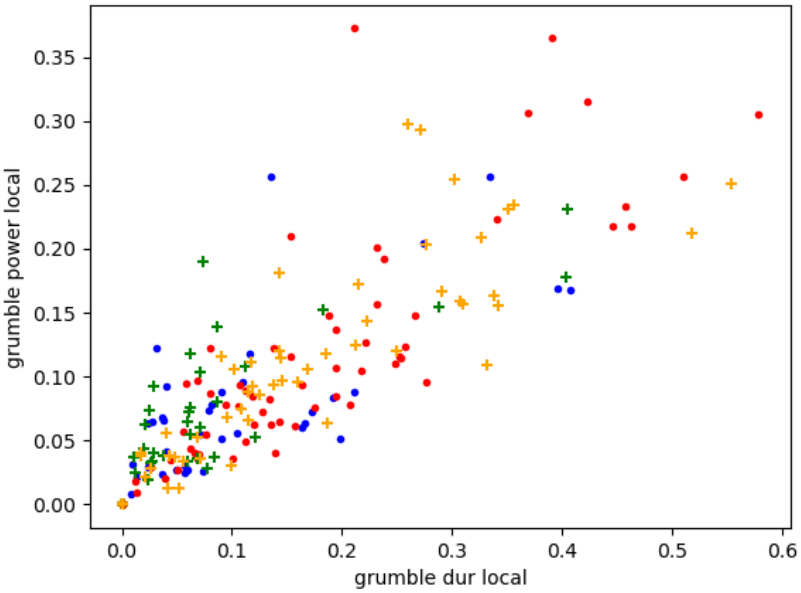

Figure 5.28 ***Scatter between grumble duration and grumble power.***

Figure 5.28 shows that grumble power is also strongly correlated to grumble duration. Grumble power is correlated to duration and can be used as a more robust measurement since calculating power needs less requirement on accurately locating the start and end time of grumble.

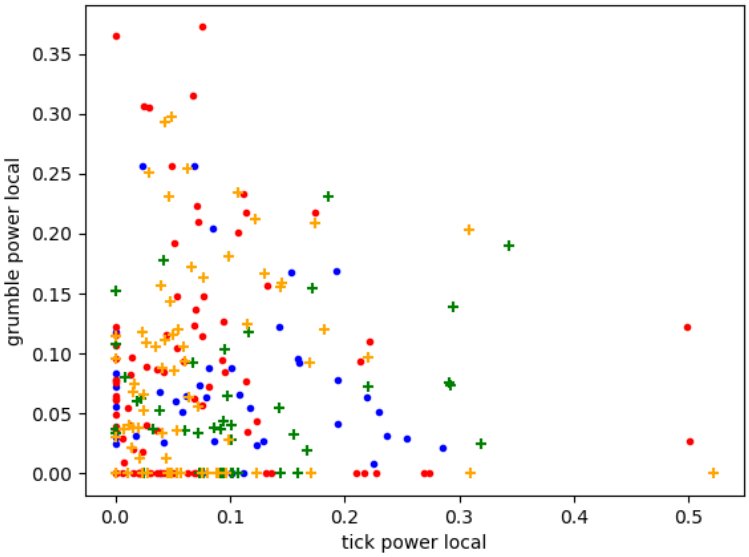

Figure 5.29 ***Scatter of grumble power and tick power: blue (healthy pre-feed), green (healthy post-feed), red (sick pre-feed), yellow (sick post-feed).***

In Figure 5.29, the scatter plot depicts the relationship between grumble power and tick power. The plot demonstrates that the healthy group exhibits lower grumble power and higher tick power in both the pre-feed and post-feed groups. However, the difference between the pre-feed and post-feed groups is not significant in the population.

5.1.6 Conclusion

This section presents a novel approach that offers affordable biosignal collection methods for integrating bowel sounds into HSNs provides valuable insights into the detection of bowel diseases in newborns, solves research question $RQ_{5.1}$ and proves hypothesis $H_{5.1}$. In this section, we address the challenge of collecting bowel sounds to detect bowel diseases in newborns within the NICU. This research aims to develop cost-effective methods for collecting and extracting important information from bowel sounds in the NICU to detect bowel diseases in newborns. Our approach involves using a miniature Bowel Sounds Sensor (BoSS) specifically designed for this purpose. These collected bowel sounds are then processed to extract relevant physiological measurements. The subsequent analysis of these measurements provides valuable information for diagnosing and monitoring bowel conditions in newborns.

The contribution of this research is the development of the BoSS, a miniature sensor designed to detect bowel diseases in newborns, thereby improving early detection within NICUs. We extract and analyse physiological measurements from bowel sounds to identify gastrointestinal issues in newborns. This study represents the first collection and analysis of bowel sounds in term newborns, addressing the need for continuous monitoring methods.

To calculate physiological measurements, we utilised the labels of ticks and grumbles for each bowel sound recording. We determined the most representative values of local physiological measurements by employing the arithmetic median of the sliding windows. This approach ensures that the local measurements offer better quality than the global measurement, as it eliminates the silent parts of the recording.

Analysis of these physiological measurements reveals that sick infants exhibit reduced tick activities and increased grumble activities, indicating potential bowel health issues. We selected local tick power and local grumble power as representative measurements of bowel activities associated with ticks and grumbles. By calculating the most representative values of these local physiological measurements using the arithmetic median of the sliding windows, we gain valuable insights into the bowel health of newborns.

5.2 Pervasive Monitoring of Gastrointestinal Health of Newborn Infants

5.2.1 Introduction

Infants needing intensive care are at risk of developing gastrointestinal problems, such as feed intolerance and NEC. NEC occurs in nearly 10% of premature infants, and accounts for 1-5% of

NICU admissions (Thompson & Bizzarro, 2008). Due to the rapid progression of NEC, early diagnosis of NEC remains challenging (Gregory, DeForge, Natale, Phillips, & Van Marter, 2011). Monitoring, early detection and prevention of bowel diseases in newborns may improve outcomes. Bowel sound monitoring provides a non-invasive clinical diagnosis (Sheu, Lin, Chen, Lee, & Lin, 2015). The absence of bowel sound often indicates NEC (Thompson & Bizzarro, 2008).

However, currently, there are no known approaches that can provide real-time, visual monitoring of bowel functions of newborn infants. Diagnosis of NEC is currently made manually using a stethoscope which is often uncomfortable and invasive for young children (Y. Huang & Song, 2019; Y. Huang et al., 2017; I. Song, 2015). Furthermore, due to the irregular pattern and occurrence of bowel sounds, locating and detecting bowel sounds from digital recordings of bowel sounds remains a challenge (Ranta et al., 2004; Sazonov et al., 2010). Manually annotating and analysing bowel sounds is time-consuming, especially when noise exists. Even doctors can only identify a few random events (Ranta et al., 2004).

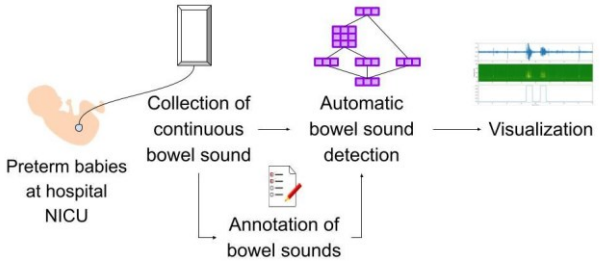

Figure 5.30 ***Overall procedure of proposed approach.***

To solve the problems, we aim to develop deep learning models that accurately segment and detect bowel sounds to improve the diagnosis of gastrointestinal problems, despite the challenges presented by the irregular pattern of bowel sounds.

Figure 5.30 illustrates how BoSS and ReLATEC are used to solve the problems. BoSS continuously records the bowel sounds of newborn infants using the micro stethoscope sensor. The bowel sound locations are then labelled for training the bowel sound detector. The labels are converted into weak labels with a lower temporal resolution because it is difficult for human listeners to identify all bowel sounds accurately due to hospital environmental noises. A ReLATEC model is then trained to detect the locations and two types of bowel sounds: short bursts and long bursts. The trained detector then provides real-time annotation of bowel sounds with locations, types, and statistics of bowel sound activities.

In our study, ReLATEC demonstrated a sensitivity of 91% and specificity of 71% in detecting short-burst bowel sounds, as well as a sensitivity of 97% and specificity of 72% in detecting long-burst bowel sounds. Despite being trained with weak labels, the model reliably identified most of the bowel sound locations labelled by human listeners and discovered new bowel sound locations.

The main contributions of our work are as follows. For the first time, we developed a miniature BoSS for preterm infants at NICUs for continuous real-time monitoring of bowel functions. A total of 171 bowel sounds of newborn infants were successfully recorded from 113 newborns at two different NICUs, and 1,267 bowel activities were successfully identified and labelled from the recordings. A new novel deep-learning approach was developed to detect the locations and types of bowel sound activities automatically.

5.2.2 Method

5.2.2.1 Data Processing

Table 5.7 ***The weak labels of short bursts generated for training and testing.***

Dataset Split	Positive	Negative	Total
Training set	1319	4731	6050
Testing set	477	1048	1525
Total	1796	5779	7575

Table 5.8 ***The weak labels of long bursts generated for training and testing.***

Dataset Split	Positive	Negative	Total
Training set	599	5451	6050
Testing set	105	1420	1525
Total	704	6871	7575

Table 5.9 ***Mean durations of bowel activities marked manually.***

Measurement	Mean Duration (seconds)
Prefeed (short burst)	0.1810
Postfeed (short burst)	0.1803
Prefeed (long burst)	0.2383
Postfeed (long burst)	0.3759

As an initial step of weak labelling, ROIs that contained bowel sounds were manually selected. Five Master of Information Technology students from James Cook University participated in the data labelling process. They received training on different types of bowel sounds, including both real and simulated bowel sounds. The first group of four students performed the initial labelling, which was then followed by cross-verification and reclassification. The remaining student later revalidated the labels. A total of 592 ROIs were selected from the 171 recordings by the first group, with each ROI containing three to four bowel activities. In total, 1,267 bowel activities were marked, including 971 for short bursts and 296 for long bursts (Table 5.7). Not all bowel activities could be identified by

human listeners due to background noises and the length of the recordings. Bowel activities often last for less than a second. Therefore, we adopted a weak labelling approach, where the labelling is not complete, but all identified bowel activities are considered correct.

To generate weak labels, we first generated positional one-hot labels using the marks. The one-hot labels were further segmented using a 0.75-second maximum pulling sliding window. Each segment was then labelled as one or zero, whereas one means there is a bowel sound at that location. This generated 7575 weak labels for each type of bowel sound. The total number of positive weak labels indicating the presence of any short bursts at the location is 1,796 and the total number of positive weak labels indicating the presence of any long bursts at the location is 704 (Table 5.8).

For training and testing, we have divided the ROIs into 10 seconds of input segments. Each input segment was then converted into a spectrogram feature data. Corresponding label data were also generated for each feature data. A total of 303 feature and label data pairs were generated: 242 for training and 61 for testing.

A single model was trained for both types of bowel sounds. For evaluation, we divided the dataset into the training dataset (80%) and the testing dataset (20%). Table 5.9 shows the mean durations of bowel activities that are manually labelled. It shows that durations of long bursts are longer than short bursts. The durations of post-feed long bursts are longer than per-feed long bursts.

5.2.2.2 Segmentation of Bowel Sound with ReLATEC

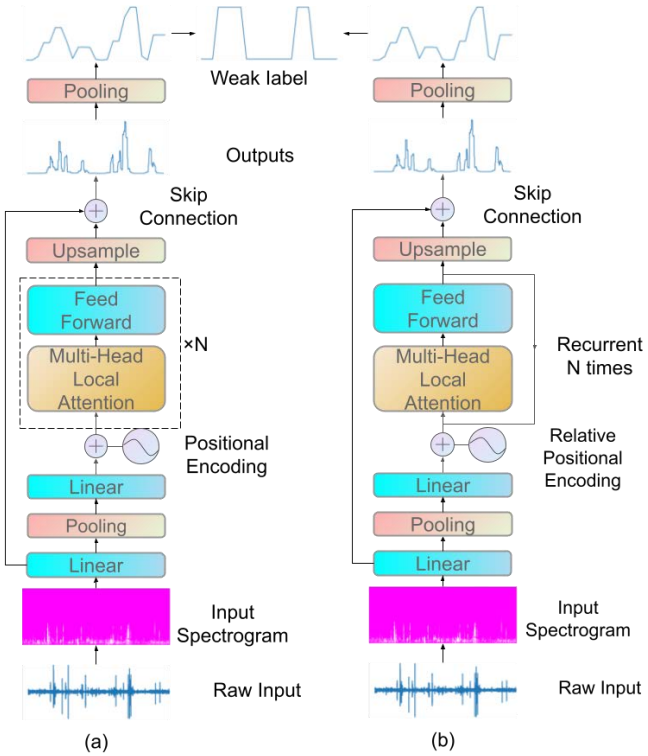

Figure 5.31 ***Comparison of (a) Vanilla Transformers and (b) ReLATEC.***

Figure 5.31 illustrates the comparison between Vanilla Transformers and ReLATEC in detecting the location and types of bowel sounds (bowel activities) from an audio stream. The input and output data have a length of 10 seconds. The input feature data is first embedded using a linear layer, followed by a pooling layer and another linear layer. Next, positional encoding is applied to encode the time locations of the feature values. The position values and the embedded feature data are then used to calculate attention. The resulting attention data is utilised to generate the output. The up sampling layer is employed to generate a 10-second length of data corresponding to the input data. The output values represent the probability of the existence of bowel activities at the respective locations in the output data.

We trained the ReLATEC model and other benchmark models using a training configuration of 50 epochs and batch sizes of 64. The models were trained using the AdamW optimizer with an initial learning rate of 0.0001 and a weight decay of 0.0001. The learning rate was adjusted using a step decay learning rate scheduler, which reduced the learning rate by 20% after each epoch. Additionally, early stopping was implemented with a patience of 5 epochs to prevent overfitting.

5.2.3 Result

5.2.3.1 Bowel Sound Segmentation

Table 5.10 ***Percentage RMSE of respiratory clinical physiological measurements.***

Performance Measurements	**Transformer**	**ReLATEC**
IoU (%)	0.621	0.770
Model Size (Megabyte)	1.950	0.124
Feed-Forward Memory Usage (Megabyte)	331.6	132.1

Table 5.10 shows the melt-down experiment. Even though ReLATEC has lower feed-forward memory usage (132.6M) than a normal vanilla Transformer (331.6M), ReLATEC has overall better segmentation performance compared to the Transformer.

5.2.3.2 Bowel Activity Location Prediction

Table 5.11 ***Performance Results of Predicting Bowel Activities Using Different Approaches.***

Approach	**Long Burst**			**Short Burst**		
	Sensitivity	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**	**Accuracy**
RNN	88.47%	**71.56%**	76.85%	90.48%	71.69%	72.98%
Transformer	85.53%	70.04%	74.89%	95.24%	72.82%	74.36%
Recurrent Transformer	88.26%	70.04%	75.74%	96.19%	**72.96%**	**74.56%**
ReLATEC	**90.99%**	70.90%	**77.18%**	**97.14%**	72.25%	73.97%

Table 5.11 shows that ReLATEC archived 90.99% sensitivity and 70.90% specificity for short bursts while it achieved 97.14% sensitivity and 72.25% specificity for long bursts.

To see whether ReLATEC performs better than existing approaches, we fixed the specificity to 70% by adjusting the decision boundary threshold and compared sensitivity values. ReLATEC achieved significantly higher sensitivity compared with existing approaches. Especially for short burst activities, it showed a 6.66% improvement. Figure 5.32 shows the ROC curves of short-burst and long-burst detections. ReLATEC achieved the highest AUC. The AUC for detecting long bursts was 93.1%, and the AUC for detecting short bursts was 88.5%. ReLATEC was superior to the conventional RNN baseline, Vanilla Transformer and recurrent Transformer.

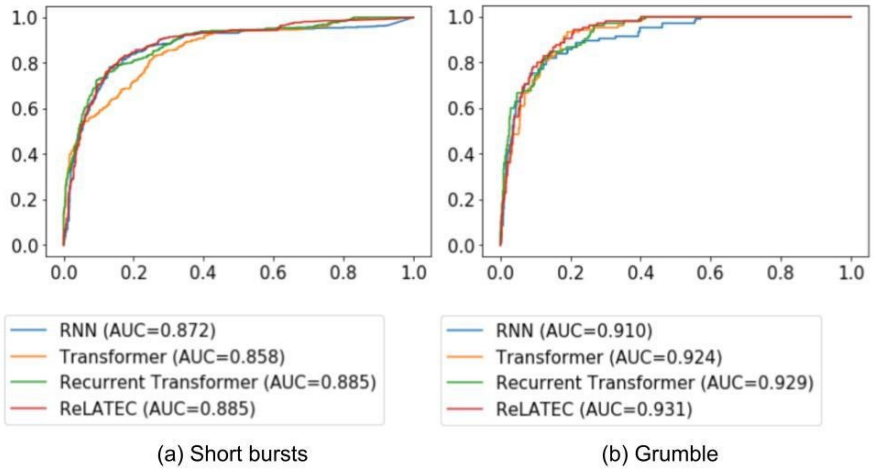

Figure 5.32 ***ROC of different models to classify (a) short burst and (b) grumble locations.***

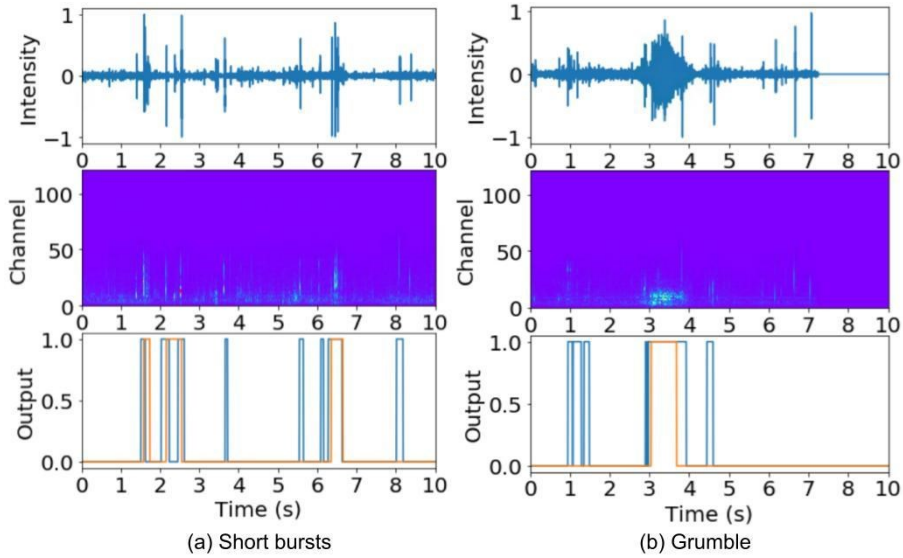

Figure 5.33 ***Segmentation visualisation of (a) short burst and (b) grumble.***

Figure 5.33 shows the raw signal, spectrogram, ground truth and prediction of short bursts and grumbles. The orange lines indicate labels by human listeners, and the blue lines indicate predictions. ReLATEC detected almost all bowel sounds (bowel activities) marked by human listeners.

Furthermore, ReLATEC also discovered and detected bowel sound activities that human listeners did not initially detect.

5.2.3.3 Clinical Statistics

Table 5.12 ***Physiological Measurements Calculated from ReLATEC Prediction.***

Measurement	**Mean Rate (per minute)**	**Std. Rate**	**Mean duration (seconds)**	**Mean gap (seconds)**
Prefeed (burst)	2.0413	1.6095	0.2130	1.7597
Postfeed (burst)	3.1754	3.8070	0.2248	1.6668
Prefeed (grumble)	1.5280	0.8269	0.2517	2.1221
Postfeed (grumble)	1.9666	0.9137	0.2955	2.1579

Table 5.12 shows physiological measurements of the bowel activities based on ReLATEC predictions. Like the measurements done using the manual labels in Table 5.9, the mean durations of long-burst bowel activities are larger than the short-burst bowel activities. The mean rates of short-bust bowel activities are higher than the post feed rates. However, due to high variance, the differences are not significant.

5.2.4 Conclusion

This section introduces a new approach that detects bowel sounds using ReLATEC and solves research question $RQ_{5.2}$ and hypothesis $H_{5.2}$. The main challenge in our study involves using deep learning models to accurately segment bowel sounds. The irregular pattern of bowel sounds poses difficulties in locating and detecting bowel activities, limiting the effectiveness of conventional deep-learning models in diagnosing gastrointestinal problems. This research aims to develop deep learning models that accurately segment and detect bowel sounds, despite the challenges posed by their irregular pattern. Our approach involves collecting bowel sounds of newborn infants using BoSS and training ReLATEC with novel weak annotations. This enables the model to accurately segment and detect bowel sounds, thereby improving the diagnosis of gastrointestinal problems in newborns. Our contribution lies in enhancing the ReLATEC model by training it with bowel sound data, resulting in improved diagnosis of gastrointestinal issues in newborns. This enhancement facilitates the extraction of meaningful information from biosignals, leading to improved healthcare decision-making processes.

The main contributions of our work are as follows: Our proposed method provides non-invasive bowel activity detection and classification. The miniature BoSS successfully recorded bowel sounds and activities in two different NICUs, demonstrating its innovation. The 3D-printed micro stethoscope was designed for comfortable and continuous monitoring of bowel functions in newborn infants. Our novel deep-learning approach, ReLATEC, significantly improves the distinction and detection of

bowel activities, even in the presence of hospital environmental noises. It provides reliable and accurate boundaries for the two types of bowel activities with high sensitivity. Furthermore, ReLATEC successfully detects bowel sound activities that human listeners initially failed to detect. The model can be trained with noisy labels and provides predictions with higher accuracy. The combination of recurrent natures such as RNN and self-attention from Transformer in ReLATEC achieves superior performance compared with RNN and Transformer alone. Despite the limitations of imbalanced data, the model produces good specificity values and robust ROC curves, with AUCs of 0.87 and 0.91 for short bursts and long bursts, respectively, indicating its reliability and suitability for medical applications.

5.3 Searching Health Information Using Bowel Sounds

5.3.1 Introduction

Gastrointestinal problems such as diarrhoea, feeding intolerance, and NEC are major health threats to term newborn babies who are often cared for in the NICU. Currently, there are no known automated bowel activity monitoring systems available to detect potential health risks of term newborn babies resulting in thousands of mortality rates per year in Australia alone.

HSNs and IoT provide solutions for remote diagnosis to enable remote diagnosis. HSNs are patient-driven communities that provide rich information to patients and caregivers (Song & Marsh, 2012). While medical professionals and offline resources are the foundation of healthcare, people still need to share knowledge on social networks. HSNs have become an important source of information for patients and physicians to make informed decisions. In recent years, many automatic recommendation systems for social networks have been developed (Song & Marsh, 2012). For example, 20% of adults have found similar health problems from others online, and most of them have benefited from HSNs (Fox, 2011). On the other hand, the IoT provides a potential solution for objectively describing health conditions with clinical measurements. IoT sensors could collect bowel sounds to provide pervasive, low-cost, and autonomous remote gastrointestinal health monitoring. Bowel sounds are non-stationary audio signals that may carry physiological information of gastrointestinal health. Previous studies performed non-invasive monitoring of intestinal sounds by extracting jitter and shimmer (K. S. Kim et al., 2011), and compared the similarity of bowel sounds (Y. Huang et al., 2017). There are also studies on long-term monitoring of gastrointestinal motility by bowel sound detection (Goto et al., 2015; Y. Huang & Song, 2019; Yin et al., 2018).

However, both HSNs and IoT have limitations. Finding relevant information remains a challenge because data from HSNs is too large to be processed manually (Alimova & Tutubalina, 2017; Miftahutdinov et al., 2017). Processing health-related information from HSNs is expensive, time-consuming, and not reliable (Achrekar et al., 2013; Alimova & Tutubalina, 2017; Cheng et al., 2017; Coppersmith et al., 2014; Elena Tutubalina & Nikolenko, 2017). In contrast, users without expertise

often encounter challenges when trying to locate relevant information. HSNs rely on users inputting keywords to index data. However, due to their lack of experience, these users may struggle with inconsistencies or errors when attempting to find appropriate search terms, which leads to misinformation.

Current automated bowel sound analysis approaches lack interpretable bowel sound pattern analysis for further understanding of bowel conditions. However, the current approaches lack interpretable bowel pattern abnormality analysis for further understanding of bowel conditions. This hinders effective searches for health-related information within HSNs, requiring manual interpretation and monitoring by trained medical professionals (Manogaran et al., 2018). Consequently, the current approach still necessitates the manual interpretation of bowel sounds and the search for relevant information on gastrointestinal health within HSNs.

To overcome these problems, we aim to convert bowel sounds into interpretable clinical measurements that represent specific physical abnormalities analysed in Section 5.1. These measurements will include characteristics such as hypoactive bowel sounds, and hyperactive bowel sounds and will serve as keywords for semantic search within HSNs. This conversion process will enhance the search capabilities of the HSNs and facilitate targeted information retrieval.

PhysioVec is a multistage deep learning framework for analysing the stream of bowel sound data for automated annotation and searching. It generates human interpretable clinical descriptions of patients from bowel sounds and uses it to automatically search HSNs. PhysioVec consists of three modules: LRT for bowel segmentation, MLI for generating clinical description, and SBERT module for the semantic search of online content.

We collected 240 bowel sounds from two NICU hospitals using an innovative miniature BoSS, that continuously monitors neonatal bowel sounds with minimal discomfort (Figure 5.1). It comprises a digital micro stethoscope, signal amplifiers specifically designed for the bowel sound frequency spectrum, and an IoT interface to send captured signals to mobile devices for recording and analysis. This non-invasive approach reduces pain for patients and does not require a trained physician (Y. Huang & Song, 2019; Y. Huang et al., 2017; I. Song, 2015).

Bowel sounds are classified into two types: ticks (short bursts) and grumbles (long bursts). We calculated clinical measurements of each bowel sound recording for each type of bowel sound. Clinical measurements are calculated in two different ways: global and local. Global measurements are calculated as the average values of the entire audio of a recording session of a patient, while local measurements are calculated as the arithmetic median of the measurements for several sliding windows of the entire audio. Local measurements showed that it is effective in removing the silent and noisy part of the record. We selected local tick power and local grumble power to represent the intestinal activity associated with ticks and grumble.

The main contributions of this method are as follows: (1) detection and segmentation of bowel sounds into two different types of bowel activities using the LRT that combines local attention and recurrent transformer to reduce overfitting and improve the performance of bowel sound segmentation; (2) extracting human interpretable physiological information through intestinal sound pattern; (3) automatically searching relevant physiological information of bowel activities with extracted physiological information on the internet. LRT achieved 0.850 and 0.792 AUC for tick and grumble segmentation. Pearson scores were 0.519 and 0.580 for estimating tick power and grumble power, respectively. PhysioVec achieved 100.00% precision in the top-one search results for bowel sounds with both vomiting and bowel obstruction. PhysioVect successfully integrated the IoT with HSNs, allowing patients and doctors to search for relevant health information directly from bowel sounds. It also provides interpretable physiological and pathological information about underlying respiratory health conditions. It can not only predict diseases but also provide patients with a wealth of information so that they can better understand their health status. Our proposed PhysioVec avoids the misinformation caused by inconsistencies or errors from the lack of expertise of users.

5.3.2 Method

5.3.2.1 Overall Process

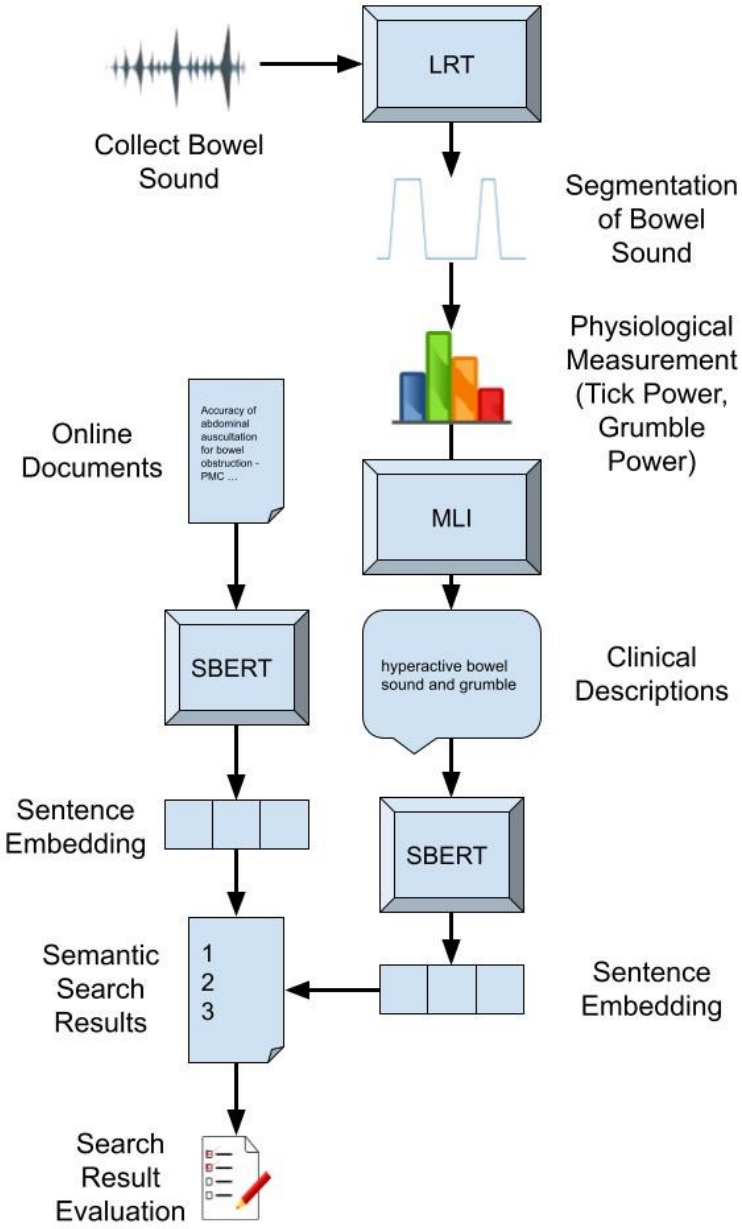

Figure 5.34 ***The overall process of PhysioVec.***

Figure 5.34 shows the overall process of the proposed PhysioVec. Bowel sounds are collected non-invasively using common IoT or digital recorders. Then the segmentation annotation of bowel sounds is generated using an LRT trained on bowel sounds with the same hyperparameters as described in Chapter 4, Section 4.2. After that, physiological measurements are extracted from segmented bowel sound. The physiological measurements are then converted into clinical descriptions by MLI based on statistics and existing knowledge. Finally, the clinical descriptions are used for searching online

documents. SBERT is applied to convert both online documents and clinical descriptions into vectors for semantic searching.

5.3.2.2 Bowel Sound Physiological Descriptions

Table 5.13 ***Healthy tick and grumble range.***

	healthy (pre-feed)	**healthy (post-feed)**
Tick Rate Local (s^{-1})	20.214 - 46.478	24.146 - 42.126
Tick Dur Local	0.052 - 0.289	0.026 - 0.360
Tick Power Local (dBm s^{-1})	0.049 - 0.214	0.032 - 0.273
Grumble Dur Local	0.000 - 0.242	0.000 -0.188
Grumble Power Local (dBm s^{-1})	0.006 -0.151	0.008 -0.143

Table 5.13 outlines the health range of bowel sound measurements for ticks and grumbles before and after feeding. In the “healthy (pre-feed)” section, low tick power is indicative of hypoactive bowel sounds, while high tick power suggests hyperactive bowel sounds. Similarly, high grumble power signifies abnormal grumbles. The “healthy (post-feed)” section follows a similar pattern, where low tick power suggests hypoactive bowel sounds, high tick power indicates hyperactive bowel sounds and high grumble power points to abnormal grumbles. These measurements provide valuable insights into the variability of bowel sound characteristics in healthy subjects, aiding in the identification and assessment of bowel sound abnormalities.

Table 5.14 ***Definitions of Bowel Rhythm Clinical Descriptions based on Bowel Physiology Measurement.***

Medical Term of Bowel Abnormality	**Tick Power**	**Grumble Power**
Hyperactive Bowel Sound	High	
Hypoactive Bowel Sound	Low	
Abnormal Bowel Sound		High

Table 5.14 presents a comprehensive elucidation of clinical descriptions of bowel rhythm abnormalities, predicated on meticulous measurements of bowel physiology. In our study, we identified three types of bowel abnormalities: hyperactive bowel sound, hypoactive bowel sound and abnormal bowel sound. The first category, hyperactive bowel sound, is characterised by an elevated tick power. This indicates a higher frequency or intensity of bowel sounds compared with those in healthy individuals. The second category, hypoactive bowel sound, is distinguished by a heightened tick power. This suggests a lower frequency or intensity of bowel sounds compared with those in healthy individuals. Lastly, the abnormal grumble sound category is manifested by higher values for grumble power.

5.3.2.3 Evaluation Metrics

We evaluate the search results with a utility function that detects whether the searched articles mention the disease name of the bowel sound. We should notice that the search queries are only descriptions of bowel sound patterns (Table 5.14). Thus, this utility function evaluates the effectiveness of the clinical descriptions since we do not train any model to detect disease names directly.

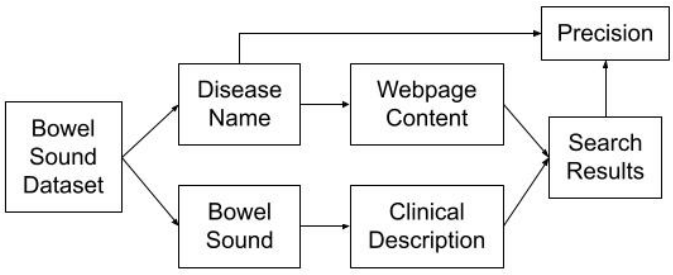

Figure 5.35 ***Process of evaluating the search results.***

Figure 5.35 shows the process of evaluating the search results. In the dataset, each bowel sound is labelled with disease names. We also collect relevant articles for each disease. The clinical description is used to search among the collected articles. Articles ranked in the top 1, top 3, top 5 and top 10 will be retrieved as search results. After that, the utility function is used to evaluate the retrieved article.

We evaluate our search results based on accuracy and recall rates. Recall rate measures the percentage of correct articles ranked *n* retrieved from all correct articles. Given that we had more than five articles for each disease, we only calculated recall rates for the first 10 results.

5.3.3 Results

Table 5.15 ***Percentage RMSE of Clinical Physiological Measurements (batch size=64, Sequence Length=300).***

Performance Measurements	**Transformer**	**LRT**	**LRT-Large**
Tick AUC	0.750	0.850	**0.851**
Grumble AUC	0.821	0.792	**0.840**
Model Size (Megabyte)	1.950	**0.141**	0.446
Feed-Forward Memory Usage (Megabyte)	4054.067	**593.739**	2591.846

Table 5.15 presents the results of the melt-down experiment. Transformer, LRT and LRT-Large were trained under the same configuration and on the same training set. LRT exhibits significantly lower feed-forward memory utilisation (0.594GB) compared with vanilla Transformer (4.054GB), and it demonstrates better overall segmentation performance. LRT-Large outperforms vanilla Transformer in both Tick and grumble detection, achieving AUC values of 0.851 and 0.840 respectively. Furthermore, LRT-Large maintains lower feed-forward memory utilisation (2.591GB).

Table 5.16 ***Pearson Correlation Coefficients of Bowel Clinical Physiological Measurements.***

Measurements	Transformer	LRT	LRT-Large
Local Tick Duration	0.462	0.543	**0.584**
Local Tick Power	0.393	0.519	**0.549**
Local Grumble Duration	**0.462**	0.387	0.418
Local Grumble Power	0.478	**0.580**	0.558

Table 5.16 presents the Pearson product-moment correlation coefficient of the proposed LRT in predicting bowel respiratory clinical physiological measurements. While LRT exhibits a lower AUC in grumble segmentation and a lower Pearson score in grumble duration estimation, it still demonstrates a superior Pearson score in grumble power estimation. The strong performance in grumble power prediction suggests that LRT focuses on the more significant and influential aspects of the grumble, which are critical for power measurement. Considering the overall good performance of LRT in both tick power and grumble power prediction, we select LRT as the model for generating physiological measurements for semantic searching.

Table 5.17 ***Precision of predicting diseases with predicted physiological measurement.***

Name of Disease	Bowel Obstruction	Vomiting
Top 1 Precision	100.00%	100.00%
Top 3 Precision	96.97%	100.00%
Top 5 Precision	92.73%	100.00%
Top 10 Precision	87.27%	100.00%
Recall Rate	62.34%	66.66%

Table 5.17 shows the accuracy of PhysioVet to find the correct literature for each bowel sound. We should also notice that bowel obstruction and vomiting are not in the training set, and we did not fine-tune the SBERT models. This prediction is based on zero-shot learning. The search results achieved 100.00% top-one precision in searching documents for bowel sounds with both vomiting and bowel obstruction, respectively. Semantic search performance suggests that most of the search results provided by PhysioVec are potentially helpful to users. The recall rates of vomiting and bowel obstruction were 72.73% and 90.91%, respectively. With a high recall rate and precision, users can get enough health information through the internet and massive data on HSNs.

5.3.4 Conclusion

This section introduces PhysioVec, a multistage deep-learning framework for indexing bowel sounds through searching HSNs directly using biosignals and solves research question $RQ_{5.3}$ and hypothesis $H_{5.3}$. In conclusion, this research addresses the challenge of enabling the semantic search of text-based contents from HSNs using bowel sounds from Health IoT. The proposed framework, PhysioVec, offers a solution by converting biosignal data into clinically interpretable descriptions that serve as keywords for semantic search within the HSNs. By utilising semantic segmentation techniques and

statistical analysis, PhysioVec extracts specific physiological measurements from bowel sounds and generates human-readable keywords for retrieving relevant health information.

The main contribution of this research is the enhancement of the ReLATEC model through training with bowel sound data, improving the diagnosis of gastrointestinal issues in newborns. This contribution enhances the understanding and extraction of meaningful information from biosignals, thereby improving healthcare decision-making processes.

The proposed framework is extendable and customisable, allowing for automatic early diagnosis and mass respiratory health monitoring in both hospital and home settings. The inclusion of visualisation of respiratory physiological measurements further aids interpretation and analysis.

However, it is important to acknowledge the limitations of this study. The generated clinical descriptions are not exhaustive, as they do not include patients' subjective descriptions and other abnormalities not covered in the mapping function. Future work could involve collecting bowel sound data in conjunction with other modalities, such as ECG, patient description text and body temperature readings, to train a model that provides more detailed and accurate descriptions of health conditions.

Additionally, the evaluation of search results is currently based on whether the retrieved articles mention the disease name. To address real-world scenarios where users require more information than just the disease name, further testing is needed to evaluate the performance of the search results.

In summary, this research presents a feasible solution that bridges health sensor-derived insights with online literature through the semantic integration of bowel-based physiological data. The framework, PhysioVec, provides an optimal first step towards intuitive respiratory monitoring by automating interpretation from captured sounds and streamlining the accessibility of relevant health information through biosignal-driven queries. Future extensions could include expanding the supported vital signs, searching clinical terminologies and patient forums, or incorporating domain adaptation techniques. Overall, this research contributes to the advancement of healthcare technology and decision-making processes in the field of respiratory Health IoT applications.

6 Conclusion and Discussion

The main contribution of this thesis is the integration of HSNs and IoT through the interpretation of biosignals to enable effective searching within HSNs. This thesis comprises a collection of published or accepted research papers. The results of the performance evaluation demonstrate that our proposed methods successfully interpret biosignals, enabling them to be utilised for effective searching within HSNs.

6.1 Summary of Contributions

This thesis aims to address the challenges in integrating HSNs with IoT devices by utilising biosignals and making them searchable in HSNs. Through a comprehensive literature review, gaps were identified in utilising various biosignals, such as ECG, breath and bowel sounds, for remote diagnosis and online medical information retrieval.

Eight research questions were formulated to guide the investigation. These questions focused on topics such as matching individuals to online support groups based on biosignals from IoT sensors, enabling semantic searching of health data in HSNs, developing deep learning models for biosignal analysis, and extracting interpretable patterns from biosignals. Corresponding hypotheses were proposed for each research question, providing a framework for testing and evaluation.

The research questions were systematically addressed in 3 technical chapters, which employed various methodologies and techniques. Key contributions of this thesis include the development of novel techniques such as Gaussian Hamming Distance, a Keyword-based Integrated HSN of Things framework, CardioVec, Recurrent Local Relation Transformer Encoder Classifier, PhysioVec, a compact Bowel Sound Sensor, and bowel sound detection models.

Evaluation results confirmed the validity of the hypotheses and demonstrated the improved performance of the developed techniques compared to existing methods. The techniques showed promising results in tasks such as disease prediction, sound segmentation, clinical description generation and information retrieval.

In conclusion, this thesis successfully addressed the research questions through rigorous experimentation and validation. The integration of HSNs and IoT through interpretable biosignals has been achieved, offering potential benefits for remote healthcare and telemedicine.

6.2 Future Work

In the conclusion section of this thesis, we have successfully demonstrated the integration of HSNN and the IoT to enhance the accessibility and accuracy of health information for patients and caregivers. Through the development of innovative frameworks and algorithms, we have made significant strides in the realm of remote healthcare and telemedicine. However, there are limitations and challenges in applying these research findings in actual health management settings that must be acknowledged and addressed.

One of the primary limitations is the reliance on large models for data interpretation and analysis. While these models have shown promise in controlled environments, their scalability and robustness in real-world scenarios, where data may be noisy or incomplete, remain a concern. Additionally, the integration of these models into existing healthcare infrastructure requires significant investment in terms of both resources and training for healthcare professionals.

Furthermore, the current research primarily focuses on the technical aspects of data collection and analysis, with less emphasis on the ethical considerations and patient privacy concerns that are integral to healthcare settings. Ensuring that the implementation of these technologies respects patient autonomy and maintains the confidentiality of sensitive health data is a challenge that must be met.

Another challenge is the generalizability of the findings across diverse populations. The research conducted herein may not be representative of all patient groups, and further studies are needed to validate the effectiveness of these models across different demographic and geographic contexts.

In terms of future work, there is a need to expand the scope of the research to include longitudinal studies that can track the long-term impact of these technologies on health outcomes. Additionally, exploring the integration of these models with EHR and other healthcare databases could provide a more comprehensive picture of patient health and inform more personalized treatment plans.

In summary, while this thesis presents a promising approach to leveraging HSNs and IoT for healthcare, it is crucial to recognize the limitations and challenges in its practical application. Future research should focus on addressing these issues, ensuring that these technologies can be effectively and responsibly integrated into the complex landscape of health management. This thesis proved that useful and interpretable physiological measurements can be extracted from biosignals for searching online health data. Our approach can be further improved for an end-to-end solution, by collecting biosignal that is associated with HSN data.

6.2.1 End-to-End Online Health Information Recommender System

Our proposed PhysioVec and other approaches are trained using existing data, which lacks biosignal data associated with online health data. This absence of online health data linked to biosignals poses a significant barrier to our study. One possible reason for this limitation could be the limited utilisation of Health IoT and HSNs at the current stage. In the future, it may be possible to develop an end-to-end recommender system by automatically collecting user actions in HSNs along with biosignals. This approach would involve training biosignals directly into word embeddings for comparison with document embeddings of SBERT, eliminating the need for extracting physiological measurements. Compared with our current approach, the end-to-end framework offers two major advantages: (1) reduced cost for identifying clinical descriptions; and (2) the potential for the model to learn to identify clinical abnormalities based on a vast amount of data once Health IoT becomes widely adopted.

6.2.2 Online Health Question Answering System

Based on the search results obtained from PhysioVec or the future recommender system, the implementation of a question-answering system becomes a viable option. An extractive question-answering system utilising a large language model such as ChatGPT can be deployed to aggregate and condense the search results. For instance, the question answering system could identify relevant diseases based on user-reported symptoms by analysing the extensive set of search results.

The potential application of the proposed PhysioVec framework in RAG tasks presents an exciting avenue for exploration. PhysioVec's capability to analyse and interpret biosignals from IoT devices could be instrumental in enhancing the quality of generated responses in RAG tasks. By utilizing the rich, clinically interpretable descriptions derived from biosignals, PhysioVec could contribute to the creation of more accurate and contextually relevant outputs in healthcare-related natural language processing tasks.

Moreover, the integration of PhysioVec with RAG could lead to the development of advanced healthcare decision support systems. These systems would be capable of not only retrieving relevant medical information but also generating new insights based on the analysis of biosignals and patient histories. This synergy between PhysioVec and RAG could significantly improve the efficiency and effectiveness of healthcare services, providing medical professionals with valuable tools for diagnosis, treatment planning, and patient education.

As we move forward, it will be essential to conduct extensive testing and validation to ensure that the integration of PhysioVec with RAG tasks meets the rigorous standards required in healthcare. This will involve iterative refinement of the models, as well as careful consideration of the ethical implications and data privacy concerns associated with handling sensitive health data. The future development of PhysioVec for RAG tasks holds promise for transforming the way healthcare information is managed and utilized, ultimately leading to improved patient outcomes and more informed clinical decision-making.

7 References

Acharya, J., & Basu, A. (2020). Deep Neural Network for Respiratory Sound Classification in Wearable Devices Enabled by Patient Specific Model Tuning. *Ieee Transactions on Biomedical Circuits and Systems, 14*(3), 535-544. doi:10.1109/TBCAS.2020.2981172

Achrekar, H., Gandhe, A., Lazarus, R., Yu, S.-H., & Liu, B. (2013, 2013//). *Online Social Networks Flu Trend Tracker: A Novel Sensory Approach to Predict Flu Trends.* Paper presented at the Biomedical Engineering Systems and Technologies, Berlin, Heidelberg.

Adame, T., Bel, A., Carreras, A., Melià-Seguí, J., Oliver, M., & Pous, R. (2018). CUIDATS: An RFID–WSN hybrid monitoring system for smart health care environments. *Future Generation Computer Systems, 78*, 602-615. doi:https://doi.org/10.1016/j.future.2016.12.023

Alimova, I., & Tutubalina, E. (2017). *Automated detection of adverse drug reactions from social media posts with machine learning.* Paper presented at the International Conference on Analysis of Images, Social Networks and Texts.

Amit, G., Gavriely, N., & Intrator, N. (2009). Cluster analysis and classification of heart sounds. *Biomedical Signal Processing and Control, 4*(1), 26-36. doi:http://dx.doi.org/10.1016/j.bspc.2008.07.003

Australian Bureau of Statistics. (2024). Provisional Mortality Statistics, Jan - Dec 2023. Retrieved from https://www.abs.gov.au/statistics/health/causes-death/provisional-mortality-statistics/latest-release#2023-deaths-in-review

Aykanat, M., Kilic, O., Kurt, B., & Saryal, S. (2017). Classification of lung sounds using convolutional neural networks. *Eurasip Journal on Image and Video Processing*, 9. doi:10.1186/s13640-017-0213-2

Azariadi, D., Tsoutsouras, V., Xydis, S., & Soudris, D. (2016, 12-14 May 2016). *ECG signal analysis and arrhythmia detection on IoT wearable medical devices.* Paper presented at the 2016 5th International Conference on Modern Circuits and Systems Technologies (MOCAST).

Bahoura, M. (2009). Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. *Computers in Biology and Medicine, 39*(9), 824-843. doi:http://dx.doi.org/10.1016/j.compbiomed.2009.06.011

Baloglu, U. B., Talo, M., Yildirim, O., Tan, R. S., & Acharya, U. R. (2019). Classification of myocardial infarction with multi-lead ECG signals and deep CNN. *Pattern Recognition Letters, 122*, 23-30. doi:10.1016/j.patrec.2019.02.016

Bardou, D., Zhang, K., & Ahmad, S. M. (2018). Lung sounds classification using convolutional neural networks. *Artificial intelligence in medicine, 88*, 58-69. doi:10.1016/j.artmed.2018.04.008

Basu, V., & Rana, S. (2020). Respiratory diseases recognition through respiratory sound with the help of deep neural network. *2020 4th International Conference on Computational Intelligence and Networks (CINE)*, 1-6.

Bokov, P., Mahut, B., Flaud, P., & Delclaux, C. (2016). Wheezing recognition algorithm using recordings of respiratory sounds at the mouth in a pediatric population. *Computers in Biology and Medicine, 70*, 40-50. doi:http://dx.doi.org/10.1016/j.compbiomed.2016.01.002

Boulos, M. N., Maramba, I., & Wheeler, S. (2006). Wikis, blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education. *BMC medical education, 6*(1), 41.

Bousseljot, R., Kreiseler, D., & Schnabel, A. (1995). Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. *Biomedizinische Technik / Biomedical Engineering $V 40*(s1), 317-318.

Bozkurt, B., Germanakis, I., & Stylianou, Y. (2018). A study of time-frequency features for CNN-based automatic heart sound classification for pathology detection. *Computers in Biology and Medicine, 100*, 132-143. doi:10.1016/j.compbiomed.2018.06.026

Charleston-Villalobos, S., Martinez-Hernandez, G., Gonzalez-Camarena, R., Chi-Lem, G., Carrillo, J. G., & Aljama-Corrales, T. (2011). Assessment of multichannel lung sounds parameterization for two-class classification in interstitial lung disease patients. *Computers in Biology and Medicine, 41*(7), 473-482. doi:http://dx.doi.org/10.1016/j.compbiomed.2011.04.009

Cheng, Q. J., Li, T. M. H., Kwok, C. L., Zhu, T. S., & Yip, P. S. F. (2017). Assessing Suicide Risk and Emotional Distress in Chinese Social Media: A Text Mining and Machine Learning Study. *Journal of medical Internet research, 19*(7), 10. doi:10.2196/jmir.7276

Coppersmith, G., Dredze, M., & Harman, C. (2014). *Quantifying mental health signals in Twitter.* Paper presented at the Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality.

Dehghani, M., Gouws, S., Vinyals, O., Uszkoreit, J., & Kaiser, Ł. (2018). Universal transformers. *arXiv preprint arXiv:1807.03819*.

Deng, S.-W., & Han, J.-Q. (2016). Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps. *Future Generation Computer Systems, 60*, 13-21. doi:http://dx.doi.org/10.1016/j.future.2016.01.010

Deperlioglu, O. (2018). Classification of Phonocardiograms with Convolutional Neural Networks. *Brain-Broad Research in Artificial Intelligence and Neuroscience, 9*(2), 22-33. Retrieved from <Go to ISI>://WOS:000435398100002

Dimoulas, C., Kalliris, G., Papanikolaou, G., Petridis, V., & Kalampakas, A. (2008). Bowel-sound pattern analysis using wavelets and neural networks with application to long-term, unsupervised, gastrointestinal motility monitoring. *Expert Systems with Applications, 34*(1), 26-41. doi:http://dx.doi.org/10.1016/j.eswa.2006.08.014

Dominguez-Morales, J. P., Jimenez-Fernandez, A. F., Dominguez-Morales, M. J., & Jimenez-Moreno, G. (2018). Deep Neural Networks for the Recognition and Classification of Heart Murmurs Using Neuromorphic Auditory Sensors. *Ieee Transactions on Biomedical Circuits and Systems, 12*(1), 24-34. doi:10.1109/tbcas.2017.2751545

Elhadad, N., Zhang, S., Driscoll, P., & Brody, S. (2014). Characterizing the Sublanguage of Online Breast Cancer Forums for Medications, Symptoms, and Emotions. *AMIA Annual Symposium Proceedings, 2014*, 516-525. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4419934/

Fox, S. (2011). *The social life of health information, 2011*: Pew Internet & American Life Project Washington, DC.

Gauy, M. M., & Finger, M. (2021). Audio MFCC-gram Transformers for respiratory insufficiency detection in COVID-19. *ArXiv, abs/2210.14085*.

Ge, J., Sun, S., Owens, J., Galvez, V., Gologorskaya, O., Lai, J. C., . . . Lai, K. (2023). Development of a Liver Disease-Specific Large Language Model Chat Interface using Retrieval Augmented Generation. *medRxiv*.

Goto, J., Matsuda, K., Harii, N., Moriguchi, T., Yanagisawa, M., & Sakata, O. (2015). Usefulness of a real-time bowel sound analysis system in patients with severe sepsis (pilot study). *Journal of Artificial Organs, 18*(1), 86-91.

Gregory, K. E., DeForge, C. E., Natale, K. M., Phillips, M., & Van Marter, L. J. (2011). Necrotizing enterocolitis in the premature infant: neonatal nursing assessment, disease pathogenesis, and clinical presentation. *Advances in neonatal care: official journal of the National Association of Neonatal Nurses, 11*(3), 155.

Griffel, B., Zia, M. K., Fridman, V., Saponieri, C., & Semmlow, J. L. (2013). Path length entropy analysis of diastolic heart sounds. *Computers in Biology and Medicine, 43*(9), 1154-1166. doi:http://dx.doi.org/10.1016/j.compbiomed.2013.05.018

Hadjileontiadis, L. J., & Rekanos, I. T. (2003). Detection of explosive lung and bowel sounds by means of fractal dimension. *Signal Processing Letters, IEEE, 10*(10), 311-314.

Hu, H., Zhang, Z., Xie, Z., & Lin, S. (2019). *Local relation networks for image recognition.* Paper presented at the Proceedings of the IEEE International Conference on Computer Vision.

Huang, Y., & Song, I. (2018). *A Better Online Method of Heart Diagnosis*. Paper presented at the Proceedings of the 3rd International Conference on Biomedical Signal and Image Processing, Seoul, Republic of Korea.

Huang, Y., & Song, I. (2019). *Indexing Biosignal for Integrated Health Social Networks.* Paper presented at the ICBBE 2019, China, Shanghai.

Huang, Y., & Song, I. (2022a, 24-26 June 2022). *Cardio Vec: Searching Heart Health Information Using ECG Signals.* Paper presented at the 2022 7th International Conference on Computational Intelligence and Applications (ICCIA).

Huang, Y., & Song, I. (2022b, 8-10 July 2022). *PhysioVec: A Multi-stage Deep-Learning Framework for Searching Online Health Information with Breath Sound.* Paper presented at the 2022 IEEE 5th International Conference on Big Data and Artificial Intelligence (BDAI).

Huang, Y., & Song, I. (2022c, 24-26 June 2022). *PhysioVec: IoT Biosignal Based Search Engine for Gastrointestinal Health.* Paper presented at the 2022 7th International Conference on Computational Intelligence and Applications (ICCIA).

Huang, Y., Song, I., Rana, P., & Koh, G. (2017, 14-19 May 2017). *Fast diagnosis of bowel activities.* Paper presented at the 2017 International Joint Conference on Neural Networks (IJCNN).

İçer, S., & Gengeç, Ş. (2014). Classification and analysis of non-stationary characteristics of crackle and rhonchus lung adventitious sounds. *Digital Signal Processing, 28*, 18-27. doi:http://dx.doi.org/10.1016/j.dsp.2014.02.001

Imran, A., Posokhova, I., Qureshi, H. N., Masood, U., Riaz, S., Ali, K., . . . Nabeel, M. (2020). AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *arXiv preprint arXiv:2004.01275*.

Jara, A. J., Zamora-Izquierdo, M. A., & Skarmeta, A. F. (2013). Interconnection Framework for mHealth and Remote Monitoring Based on the Internet of Things. *IEEE Journal on Selected Areas in Communications, 31*(9), 47-65. doi:10.1109/JSAC.2013.SUP.0513005

Jin, M., Yu, Q., Zhang, C., Shu, D., Zhu, S., Du, M., . . . Meng, Y. (2024). Health-LLM: Personalized Retrieval-Augmented Disease Prediction Model. *arXiv preprint arXiv:2402.00746*.

Kamel Boulos, M. N., & Wheeler, S. (2007). The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education1. *Health Information & Libraries Journal, 24*(1), 2-23.

Kang, S. H., Joe, B., Yoon, Y., Cho, G. Y., Shin, I., & Suh, J. W. (2018). Cardiac Auscultation Using Smartphones: Pilot Study. *Jmir Mhealth and Uhealth, 6*(2), 11. doi:10.2196/mhealth.8946

Kim, J., & Min, M. (2024). From RAG to QA-RAG: Integrating Generative AI for Pharmaceutical Regulatory Compliance Process. *arXiv preprint arXiv:2402.01717*.

Kim, K. S., Seo, J. H., Ryu, S. H., Kim, M. H., & Song, C. G. (2011). Estimation algorithm of the bowel motility based on regression analysis of the jitter and shimmer of bowel sounds. *Computer Methods and Programs in Biomedicine, 104*(3), 426-434. doi:10.1016/j.cmpb.2011.02.014

Kochetov, K., Putin, E., Azizov, S., Skorobogatov, I., & Filchenkov, A. (2017). Wheeze Detection Using Convolutional Neural Networks. In E. Oliveira, J. Gama, Z. Vale, & H. L. Cardoso (Eds.), *Progress in Artificial Intelligence* (Vol. 10423, pp. 162-173). Cham: Springer International Publishing Ag.

Kucharski, D., Grochala, D., Kajor, M., & Kantoch, E. (2018). A Deep Learning Approach for Valve Defect Recognition in Heart Acoustic Signal. In L. Borzemski, J. Swiatek, & Z. Wilimowska (Eds.), *Information Systems Architecture and Technology, Pt I* (Vol. 655, pp. 3-14). Cham: Springer International Publishing Ag.

Kumar, P. M., & Gandhi, U. D. (2018). A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases. *Computers & Electrical Engineering, 65*, 222-235.

Laguarta, J., Hueto, F., & Subirana, B. (2020). COVID-19 Artificial Intelligence Diagnosis using only Cough Recordings. *IEEE Open Journal of Engineering in Medicine and Biology*, 1-1. doi:10.1109/OJEMB.2020.3026928

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.

Lei, B., Rahman, S. A., & Song, I. (2014). Content-based classification of breath sound with enhanced features. *Neurocomputing, 141*, 139-147.

Lei, Y., Chungui, L., & Sen, T. (2011, 17-18 Oct. 2011). *Community Medical Network (CMN): Architecture and implementation.* Paper presented at the 2011 Global Mobile Congress.

Limam, M., Precioso, F., & Ieee. (2017). Atrial Fibrillation Detection and ECG Classification based on Convolutional Recurrent Neural Network. In *2017 Computing in Cardiology*. Los Alamitos: Ieee Computer Soc.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., . . . Guo, B. (2021). *Swin transformer: Hierarchical vision transformer using shifted windows.* Paper presented at the Proceedings of the IEEE/CVF international conference on computer vision.

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., . . . Chen, S. H. A. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods, 53*(4), 1689-1696. doi:10.3758/s13428-020-01516-y

Mamun, K. A. A., & McFarlane, N. (2016). Integrated real time bowel sound detector for artificial pancreas systems. *Sensing and Bio-Sensing Research, 7*, 84-89. doi:http://dx.doi.org/10.1016/j.sbsr.2016.01.004

Manogaran, G., Varatharajan, R., Lopez, D., Kumar, P. M., Sundarasekar, R., & Thota, C. (2018). A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system. *Future Generation Computer Systems, 82*, 375-387.

Marshall, A., & Boussakta, S. (2007). Signal analysis of medical acoustic sounds with applications to chest medicine. *Journal of the Franklin Institute, 344*(3–4), 230-242. doi:http://dx.doi.org/10.1016/j.jfranklin.2006.08.003

Mdhaffar, A., Chaari, T., Larbi, K., Jmaiel, M., & Freisleben, B. (2017, 6-8 July 2017). *IoT-based health monitoring via LoRaWAN.* Paper presented at the IEEE EUROCON 2017 -17th International Conference on Smart Technologies.

Mehari, T., & Strodthoff, N. (2022). Self-supervised representation learning from 12-lead ECG data. *Computers in Biology and Medicine, 141*, 105114.

Miao, J., Thongprayoon, C., Suppadungsuk, S., Garcia Valencia, O. A., & Cheungpasitporn, W. (2024). Integrating Retrieval-Augmented Generation with Large Language Models in Nephrology: Advancing Practical Applications. *Medicina, 60*(3), 445.

Miftahutdinov, Z., Tropsha, A., & Tutubalina, E. (2017). *Identifying disease-related expressions in reviews using conditional random fields*.

Mihalcea, R., & Tarau, P. (2004). *Textrank: Bringing order into text.* Paper presented at the Proceedings of the 2004 conference on empirical methods in natural language processing.

Miller, E. A., & Pole, A. (2010). Diagnosis blog: checking up on health blogs in the blogosphere. *American Journal of Public Health, 100*(8), 1514-1519.

Moummad, I., & Farrugia, N. (2022). Supervised Contrastive Learning for Respiratory Sound Classification. *ArXiv, abs/2210.16192*.

Mozaffarian, D., Benjamin, E., Go, A., Arnett, D., Blaha, M., Cushman, M., . . . Fullerton, H. (2016). Heart disease and stroke statistics-2016 update: a report from the American Heart Association. *Circulation, 133*(4), e38.

Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., . . . Howard, V. J. (2015). Heart disease and stroke statistics--2015 update: A report from the American Heart Association. *Circulation, 131*(4), e29.

Mukherjee, S., Gamble, P., Ausin, M. S., Kant, N., Aggarwal, K., Manjunath, N., . . . Busacca, S. (2024). Polaris: A Safety-focused LLM Constellation Architecture for Healthcare. *arXiv preprint arXiv:2403.13313*.

Na, J.-C., Kyaing, W. Y. M., Khoo, C. S., Foo, S., Chang, Y.-K., & Theng, Y.-L. (2012). *Sentiment classification of drug reviews using a rule-based linguistic approach.* Paper presented at the International conference on asian digital libraries.

Naves, R., Barbosa, B. H. G., & Ferreira, D. D. (2016). Classification of lung sounds using higher-order statistics: A divide-and-conquer approach. *Computer Methods and Programs in Biomedicine, 129*, 12-20. doi:http://dx.doi.org/10.1016/j.cmpb.2016.02.013

Nilanon, T., Yao, J., Hao, J., Purushotham, S., & Liu, Y. (2016). *Normal/abnormal heart sound recordings classification using convolutional neural network.* Paper presented at the Computing in Cardiology Conference (CinC), 2016.

Pan, J., & Tompkins, W. J. (1985). A Real-Time QRS Detection Algorithm. *IEEE Transactions on Biomedical Engineering, BME-32*(3), 230-236. doi:10.1109/TBME.1985.325532

Potes, C., Parvaneh, S., Rahman, A., & Conroy, B. (2016). *Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds.* Paper presented at the Computing in Cardiology Conference (CinC), 2016.

Puustjärvi, J., & Puustjärvi, L. (2011, 11-13 May 2011). *Automating remote monitoring and information therapy: An opportunity to practice telemedicine in developing countries.* Paper presented at the 2011 IST-Africa Conference Proceedings.

Rahmani, A. M., Gia, T. N., Negash, B., Anzanpour, A., Azimi, I., Jiang, M., & Liljeberg, P. (2018). Exploiting smart e-Health gateways at the edge of healthcare Internet-of-Things: A fog computing approach. *Future Generation Computer Systems, 78*, 641-658. doi:https://doi.org/10.1016/j.future.2017.02.014

Ranta, R., Louis-Dorr, V., Heinrich, C., Wolf, D., & Guillemin, F. (2004). *Principal component analysis and interpretation of bowel sounds.* Paper presented at the Proceeding of the 26th Annual International Conference of the IEEE EMBS.

Rasid, M., Musa, W., Kadir, N., Noor, A., Touati, F., Mehmood, W., . . . Mnaouer, A. B. (2014). *Embedded gateway services for Internet of Things applications in ubiquitous healthcare.* Paper presented at the Information and Communication Technology (ICoICT), 2014 2nd International Conference on.

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Rubin, J., Abreu, R., Ganguli, A., Nelaturi, S., Matei, I., & Sricharan, K. (2016). *Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients.* Paper presented at the Computing in Cardiology Conference (CinC), 2016.

Ruuskanen, O., Lahti, E., Jennings, L. C., & Murdoch, D. R. (2011). Viral pneumonia. *The Lancet, 377*(9773), 1264-1275. doi:10.1016/S0140-6736(10)61459-6

Sánchez Morillo, D., Astorga Moreno, S., Fernández Granero, M. Á., & León Jiménez, A. (2013). Computerized analysis of respiratory sounds during COPD exacerbations. *Computers in Biology and Medicine, 43*(7), 914-921. doi:http://dx.doi.org/10.1016/j.compbiomed.2013.03.011

Sazonov, E. S., Makeyev, O., Schuckers, S., Lopez-Meyer, P., Melanson, E. L., & Neuman, M. R. (2010). Automatic detection of swallowing events by acoustical means for applications of monitoring of ingestive behavior. *Biomedical Engineering, IEEE Transactions on, 57*(3), 626-633.

Sengupta, N., Sahidullah, M., & Saha, G. (2016). Lung sound classification using cepstral-based statistical features. *Computers in Biology and Medicine, 75*, 118-129. doi:http://dx.doi.org/10.1016/j.compbiomed.2016.05.013

Sharif, H., Zaffar, F., Abbasi, A., & Zimbra, D. (2014). *Detecting adverse drug reactions using a sentiment classification framework.*

Sheu, M. J., Lin, P. Y., Chen, J. Y., Lee, C. C., & Lin, B. S. (2015). Higher-Order-Statistics-Based Fractal Dimension for Noisy Bowel Sound Detection. *Ieee Signal Processing Letters, 22*(7), 789-793. doi:10.1109/lsp.2014.2369856

Song, I. (2015, 12-17 July 2015). *Diagnosis of pneumonia from sounds collected using low cost cell phones.* Paper presented at the 2015 International Joint Conference on Neural Networks (IJCNN).

Song, I. (2015). *Gaussian Hamming Distance.* Paper presented at the International Conference on Neural Information Processing.

Song, I., Huang, Y., Koh, T. H. H. G., & Rajadurai, V. S. (2021, 2021//). *Pervasive Monitoring of Gastrointestinal Health of Newborn Babies.* Paper presented at the PRICAI 2021: Trends in Artificial Intelligence, Cham.

Song, I., & Marsh, N. V. (2012). Anonymous indexing of health conditions for a similarity measure. *Information Technology in Biomedicine, IEEE Transactions on, 16*(4), 737-744.

Subirana, B., Hueto, F., Rajasekaran, P., Laguarta, J., Puig, S., Malvehy, J., . . . Valle, J. F. M. (2020). Hi Sigma, do I have the Coronavirus?: Call for a New Artificial Intelligence Approach to Support Health Care Professionals Dealing With The COVID-19 Pandemic. *arXiv preprint arXiv:2004.06510*.

Sun, X., Ye, J., & Ren, F. (2015). *Hybrid Model Based Influenza Detection with Sentiment Analysis from Social Networks.* Paper presented at the Chinese National Conference on Social Media Processing.

Taplidou, S. A., & Hadjileontiadis, L. J. (2007a). Nonlinear analysis of wheezes using wavelet bicoherence. *Computers in Biology and Medicine, 37*(4), 563-570. doi:http://dx.doi.org/10.1016/j.compbiomed.2006.08.007

Taplidou, S. A., & Hadjileontiadis, L. J. (2007b). Wheeze detection based on time-frequency analysis of breath sounds. *Computers in Biology and Medicine, 37*(8), 1073-1083. doi:http://dx.doi.org/10.1016/j.compbiomed.2006.09.007

Tejeda, R. G., & Falcon, M. C. (2019). Optimization of Algorithms for Real-Time ECG Beats Classification. In L. Lhotska, L. Sukupova, I. Lackovic, & G. S. Ibbott (Eds.), *World Congress on Medical Physics and Biomedical Engineering 2018, Vol 2* (Vol. 68, pp. 335-338). New York: Springer.

Thompson, A. M., & Bizzarro, M. J. (2008). Necrotizing enterocolitis in newborns: pathogenesis, prevention and management. *Drugs, 68*, 1227-1238.

Tutubalina, E., Miftahutdinov, Z., Nikolenko, S., & Malykh, V. (2018). Medical concept normalization in social media posts with recurrent neural networks. *Journal of Biomedical Informatics, 84*, 93-102. doi:https://doi.org/10.1016/j.jbi.2018.06.006

Tutubalina, E., & Nikolenko, S. (2017). Combination of deep recurrent neural networks and conditional random fields for extracting adverse drug reactions from user reviews. *Journal of Healthcare Engineering, 2017*.

Tutubalina, E., & Nikolenko, S. (2018). Exploring convolutional neural networks and topic models for user profiling from drug reviews. *Multimedia Tools and Applications, 77*(4), 4791-4809. doi:10.1007/s11042-017-5336-z

UNICEF, & Diarrhoea, W. (2012). Why children are still dying and what can be done, 2009. In.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). *Attention is all you need.* Paper presented at the Advances in neural information processing systems.

Verma, P., & Sood, S. K. (2018). Cloud-centric IoT based disease diagnosis healthcare framework. *Journal of Parallel and Distributed Computing, 116*, 27-38.

Vong, J., & Song, I. (2015). Automated Health Care Services. In *Emerging Technologies for Emerging Markets* (pp. 89-102): Springer.

Wang, F., Syeda-Mahmood, T., & Beymer, D. (2007). *Finding disease similarity by combining ECG with heart auscultation sound.* Paper presented at the 2007 Computers in Cardiology.

Wang, J., Yang, Z., Yao, Z., & Yu, H. (2024). JMLR: Joint Medical LLM and Retrieval Training for Enhancing Reasoning and Professional Question Answering Capability. *arXiv preprint arXiv:2402.17887*.

Wang, X., Zhao, K., & Street, N. (2014). *Social support and user engagement in online health communities.* Paper presented at the International Conference on Smart Health.

Wang, X., Zhao, K., & Street, N. (2017). Analyzing and Predicting User Participations in Online Health Communities: A Social Support Perspective. *Journal of medical Internet research, 19*(4), e130. doi:10.2196/jmir.6834

White, L., Togneri, R., Liu, W., & Bennamoun, M. (2016). *Generating bags of words from the sums of their word embeddings.* Paper presented at the International Conference on Intelligent Text Processing and Computational Linguistics.

Wiesner, M., & Pfeifer, D. (2014). Health recommender systems: concepts, requirements, technical basics and challenges. *International journal of environmental research and public health, 11*(3), 2580-2607.

World Health Organization. (2023). World health statistics 2023: monitoring health for the SDGs, sustainable development goals. In: Geneva: World Health Organization. Licence: CC BY-NC-SA.

Xiong, Z. H., Nash, M. P., Cheng, E., Fedorov, V. V., Stiles, M. K., & Zhao, J. C. (2018). ECG signal classification for the detection of cardiac arrhythmias using a convolutional recurrent neural network. *Physiological Measurement, 39*(9), 10. doi:10.1088/1361-6579/aad9ed

Xu, J., Murphy, S. L., Kochanek, K. D., & Arias, E. (2022). Mortality in the United States, 2021. *NCHS Data Brief*(456), 1-8.

Yang, G., Xie, L., Mäntysalo, M., Zhou, X., Pang, Z., Da Xu, L., . . . Zheng, L.-R. (2014). A health-IoT platform based on the integration of intelligent packaging, unobtrusive bio-sensor, and intelligent medicine box. *IEEE transactions on industrial informatics, 10*(4), 2180-2191.

Yildirim, O. (2018). A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. *Computers in Biology and Medicine, 96*, 189-202. doi:10.1016/j.compbiomed.2018.03.016

Yin, Y., Jiang, H., Feng, S., Liu, J., Chen, P., Zhu, B., & Wang, Z. (2018). Bowel sound recognition using SVM classification in a wearable health monitoring system. *Science China Information Sciences, 61*(8), 084301. doi:10.1007/s11432-018-9395-5

Zhang, L., Hall, M., & Bastola, D. (2018). Utilizing Twitter data for analysis of chemotherapy. *International journal of medical informatics, 120*, 92-100. doi:https://doi.org/10.1016/j.ijmedinf.2018.10.002

Zhang, S., Grave, E., Sklar, E., & Elhadad, N. (2017). Longitudinal analysis of discussion topics in an online breast cancer community using convolutional neural networks. *Journal of Biomedical Informatics, 69*, 1-9. doi:https://doi.org/10.1016/j.jbi.2017.03.012

Zhang, W. J., Han, J. Q., & Ieee. (2017). Towards Heart Sound Classification Without Segmentation Using Convolutional Neural Network. In *2017 Computing in Cardiology*. Los Alamitos: Ieee Computer Soc.

Zhao, K., Yen, J., Greer, G., Qiu, B., Mitra, P., & Portier, K. (2014). Finding influential users of online health communities: a new metric based on sentiment influence. *Journal of the American Medical Informatics Association, 21*(e2), e212-e218.

Zhu, Y., Ren, C., Xie, S., Liu, S., Ji, H., Wang, Z., . . . Zhu, X. (2024). REALM: RAG-Driven Enhancement of Multimodal Electronic Health Records Analysis via Large Language Models. *arXiv preprint arXiv:2402.07016*.