DEVELOPING A PHYSICIAN-PATIENT SPEECH RECOGNITION AND DOCUMENTATION MODEL FOR NIGERIAN HOSPITALS

ETUKUDO, DEBORAH BASSEY
(20PCH02191)

DECEMBER, 2022

DEVELOPING A PHYSICIAN-PATIENT SPEECH RECOGNITION AND DOCUMENTATION MODEL FOR NIGERIAN HOSPITALS

BY

ETUKUDO, DEBORAH BASSEY
(20PCH02191)
B.Sc. Computer Science, Landmark University, Omu-Aran

A DISSERTATION SUBMITTED TO THE SCHOOL OF POSTGRADUATE STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE (M.Sc.) DEGREE IN MANAGEMENT INFORMATION SYSTEM IN THE DEPARTMENT OF COMPUTER AND INFORMATION SCIENCES, COLLEGE OF SCIENCE AND TECHNOLOGY, COVENANT UNIVERSITY, OTA, OGUN STATE, NIGERIA

DECEMBER, 2022

ACCEPTANCE

This is to attest that this dissertation is accepted in partial fulfillment of the requirements for the award of the degree of MASTER of Sciences in Management Information System in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Nigeria.

Miss. Adefunke F. Oyinloye
(Secretary, School of Postgraduate Studies) **Signature and Date**

Prof Akan B. Williams
(Dean, School of Postgraduate Studies) **Signature and Date**

DECLARATION

I, **ETUKUDO, DEBORAH BASSEY (20PCH02191),** declare that this research was carried out by me under the supervision of Dr. Olamma U. Iheanetu of the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Nigeria. I attest that the dissertation has not been presented either wholly or partially for the award of any degree elsewhere. All sources of data and scholarly information used in this dissertation are duly acknowledged.

ETUKUDO, DEBORAH BASSEY

Signature and Date

CERTIFICATION

We certify that this dissertation titled **"DEVELOPING A PHYSICIAN-PATIENT SPEECH RECOGNITION AND DOCUMENTATION MODEL FOR NIGERIAN HOSPITALS"** is an original research work carried out by **ETUKUDO, DEBORAH BASSEY (20PCH02191),** in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Nigeria under the supervision of Dr. Olamma U. Iheanetu. We have examined and found this work acceptable as part of the requirements for the award of Master of Science in Management Information System.

Dr. Olamma U. Iheanetu
(Supervisor) **Signature and Date**

Prof. Olufunke O. Oladipupo
(Head of Department) **Signature and Date**

Prof. Olusegun Folorunso
(External Examiner) **Signature and Date**

Prof. Akan B. Williams
(Dean, School of Postgraduate Studies) **Signature and Date**

DEDICATION

I dedicate this work to the Almighty God, for His infinite wisdom, grace, and love over my life. Also, this work is dedicated to my loving parents who have both worked exceptionally hard to set me up for success.

ACKNOWLEDGEMENTS

I appreciate God almighty for the strength, grace, wisdom, and understanding to execute this study. Many thanks to the head of the department, Prof. Olufunke Oladipupo, and the entire members of the faculty for this opportunity. I appreciate my supervisor, Dr. Olamma U. Iheanetu for her guidance, support, and encouragement. I also appreciate my family, friends, and colleagues for their help and support.

TABLE OF CONTENTS

CONTENTS		PAGES
COVER PAGE		**i**
TITLE PAGE		**ii**
DECLARATION		**iv**
CERTIFICATION		**v**
DEDICATION		**vi**
ACKNOWLEDGEMENTS		**vii**
TABLE OF CONTENTS		**viii**
LIST OF FIGURES		**xi**
LIST OF TABLES		**xiii**
LIST OF ABBREVIATIONS		**xiv**
ABSTRACT		**xvi**
CHAPTER ONE: INTRODUCTION		**1**
1.1	Background to the Study	1
1.2	Statement of the Problem	5
1.3	Aim and Objectives of the Study	6
1.4	Research Methodology	6
1.5	Significance of the Study	7
1.6	Scope and Limitations of the Study	7
1.7	Organisation of the Dissertation	8
CHAPTER TWO: LITERATURE REVIEW		**9**
2.1	Introduction	9
2.2	The Evolution of the Electronic Health Records	9
2.3	SOAP Medical Keeping	10
	2.3.1 Subjective	11
	2.3.2 Objective	11
	2.3.3 Assessment	11
	2.3.4 Plan	12
2.4	Information Extraction	12
	2.4.1 State of the Art Information Extraction Techniques (overview)	12
	2.4.2 Audio Information Extraction	13
	2.4.3 Noteworthy Utterance	17
2.5	Machine Learning Approaches to ASR Development	17

2.5.1 Types of Machine Learning 19
2.6 Deep Learning Approaches to ASR Development 21
2.6.1 Classes of Deep Learning Model 22
2.6.2 Deep Learning Algorithm 23
2.6.3 Advantages of Deep Learning Algorithm 24
2.7 Context of Unstructured Data and Structured Data 25
2.8 Challenges in Analysis of Electronic Health Record Data 27
2.9 Data in Automatic Speech Recognition 31
2.10 Confusion Matrix 34
2.11 Review of Related Works 37
CHAPTER THREE: METHODOLOGY 39
3.1 Introduction 39
3.2 Survey 41
3.3 Existing Model (Wav2vec) 43
3.4 The Proposed Model 44
3.5 Data Collection 47
3.5.1 Speech Data Collection 47
3.5.2 Text Classification Collection 49
3.6 Data Pre-processing 50
3.6.1 Pre-processing for Text Classification 50
3.6.2 Audio Dataset 51
3.7 The Subsystems 52
3.7.1 Speech Recognition System 52
3.7.2 Text Summarization 53
3.7.3 Text Classification 56
3.8 Evaluation 58
3.8.1 Evaluation metrics 58
3.8.2 Dataset used for Evaluation 60
3.9 Development Environment 60
CHAPTER FOUR: RESULTS AND DISCUSSION 62

4.1 Introduction 62
4.2 Data Collection and Pre-processing 62
4.2.1 Speech Data Pre-processing 62
4.2.2 Text Data Pre-processing 64
4.3 Data Visualization 67
4.4 Evaluation 70
4.4.1 Speech Recognition Result 70
4.4.2 Text Classification Result 73
4.5 Simulation and Testing 79
4.5.1 Web Application Implementation 80
4.6 Discussion 82
CHAPTER FIVE: CONCLUSION AND RECOMMENDATION 87
5.1 Summary 87
5.2 Conclusion 87
5.3 Contributions to Knowledge 88
5.4 Recommendations 88
REFERENCES 89

LIST OF FIGURES

FIGURES TITLE OF FIGURES PAGES

Figure 2.1: The official Google Cloud Speech-to-Text website, with the product's demonstration interface in the foreground. 33

Figure 2.2: Structure of Confusion Metrics 34

Figure 3.1: Methodological flow diagram 40

Figure 3.2: Gender distribution of respondent 41

Figure 3.3: Respondent hospital type 42

Figure 3.4: Average time duration for consultation 42

Figure 3.5: Number of Patients they attend to Daily 42

Figure 3.6: Doctors Years of Experience 43

Figure 3.7: How Nigerian hospitals document consultation 43

Figure 3.8: Architecture of existing Wav2vec model 44

Figure 3.9: Proposed model architecture 45

Figure 3.10: Activity diagram of the flow process 46

Figure 3.11: Overview of the procedure for speech data collection 47

Figure 3.12: Audio labeling using audacity 51

Figure 3.13: Speech recognition architecture 53

Figure 3.14: Extractive Summarization Process 53

Figure 3.15: Screenshot of spacy pre-processing 54

Figure 3.16: Screenshot of Spacy Summarization 55

Figure 3.17: MLP Architecture Overview 57

Figure 3.18: Screenshot of a text classification model 58

Figure 3.19: Screenshot of system specification 61

Figure 4.1: Screenshot of some of the audio data collected 63

Figure 4.2: Screenshot of some the audio data labeled 63

Figure 4.3: Before noise was removed from the audio 64

Figure 4.4: After noise was removed from the audio 64

Figure 4.5: Screenshot of some labeled SOAP notes. 65

Figure 4.6: Screenshot of the data pre-processing 65

Figure 4.7: Screenshot of one hot encoding in soap note 66
Figure 4.8: Splitting of the dataset for SOAP classification 66
Figure 4.9: Data visualization of the distributed SOAP notes. 67
Figure 4.10: Data visualization of the SOAP note. 68
Figure 4.11: Data visualization of the subjective class 68
Figure 4.12: Data visualization of the objective class 69
Figure 4.13: Data visualization of the assessment class 69
Figure 4.14: Data visualization of the plan class 70
Figure 4.15: Confusion metrics for the machine learning techniques 76
Figure 4.16: Confusion metrics for MLP technique 78
Figure 4.17: Screenshot of the developed web application for the speech recognition system. 80
Figure 4.18: Screenshot of the developed web application for the SOAP system (i) 81
Figure 4.19: Screenshot of the developed web application for the SOAP system (ii 81
Figure 4.20: Number of times system was used for testing 83
Figure 4.21: Reason for using the system 83
Figure 4.22: Screenshot of the respondent responses concerning the system usability 84
Figure 4.23: Screenshot of the respondent responses concerning the system user interface. 84
Figure 4.24: Screenshot of the respondent responses concerning how well the app recognized the consultations. 84
Figure 4.25: Screenshot of the respondent responses concerning how long consultation took using the app. 85
Figure 4.26: Feedback gotten from system testing 85

LIST OF TABLES

TABLES	TITLE OF TABLES	PAGES
Table 1.1:	Summary of objectives	6
Table 3.1:	The medical conversation scripts	48
Table 3.2:	Table showing the counts of the labeled and manually labeled conversation in SOAP	51
Table 3.3:	Evaluation metrics (i)	59
Table 3.4:	Evaluation metrics (ii)	60
Table 4.1:	Overall result from ASR evaluation	70
Table 4.2:	Finetuned Wav2vec model trained on the primary dataset	72
Table 4.3:	Wav2vec model trained on TIMIT dataset	72
Table 4.4:	Machine learning text classification evaluation result	74
Table 4.5:	Evaluation Metrics table of the SOAP classification	76
Table 4.6:	Confusion metrics table of the SOAP classification	78
Table 4.7:	Result from text classification	79

LIST OF ABBREVIATIONS

AED	Acoustic Event Detection
AI	Artificial Intelligence
ANN	Artificial Neural Network
ASR	Automatic Speech Recognition
ATM	Amazon Transcribe Medical
BP	Back Propagation
CNN	Convolutional Neural Network
DBN	Deep Belief Networks
DL	Deep Learning
DNN	Deep Neural Network
DT	Decision Tree
DQA	Data Quality Assessment
EHR	Electronic Health Record
EMR	Electronic Medical Record
GP	Gaussian Process
GPU	Graphics Processing Unit
FF	Forward Propagation
HMM	Hidden Markov Model
IE	Information Extraction
LSTM	Long Short-Term Memory
ML	Machine Learning
MLP	Multilayer Perceptron
NLP	Natural Language Processing
NLG	Natural Language Generation
NN	Neural Network
RBM	Restricted Boltzmann Machine

RNN	Recurrent Neural Network
SR	Speech Recognition
SOAP	Subjective, Objective, Assessment, Plan
SVM	Support Vector Machine
TIMIT	Texas Instruments/Massachusetts Institute of Technology.

ABSTRACT

Despite the increase in the use of automated systems for physician-patient consultation in developed countries, Nigerian hospitals are still actively using the traditional approach which implies using pen and paper or typing on the computer in documenting and extracting the conversation between the physician and the patient. This has been observed to cause long consultation time, burnout during consultations, bad consultation experience for the patient, etc. Due to the Nigerian accent, the present speech recognition system when tested with a Nigerian accent could not properly detect it. This study aims to develop an automatic speech recognition system for extracting and classifying structured data from physician-patient conversations done in a Nigerian hospital. A pretrained Wav2vec model was retrained with the primary data used in this work for better detection of the Nigerian accent. The accuracy score of the SOAP classification using a deep learning approach, gave a total of 83%, compared to that of the machine learning approach which gave a total accuracy of 82%. For the speech recognition system evaluation, the Word error rate, word information lost, and match error rate gave a better accuracy of 0.889, 0.800, and 0.896 respectively as compared to the TIMIT dataset which gave 0.999, 0.999, and 0.999 respectively. The proposed system performs better in the identification of the Nigerian accent than the existing system.

Keywords: Automatic Speech Recognition, Wav2vec, Electronic Health Record, SOAP Notes, Text Classification, MLP.