UNIVERSITI TEKNIKAL MALAYSIA MELAKA

NEURAL NETWORK-BASED ECHOCARDIOGRAM VIDEO CLASSIFICATION BY INCORPORATING DYNAMIC INFORMATION AND ATTENTION MECHANISM

YE ZI

DOCTOR OF PHILOSOPHY

2022

Faculty of Information and Communication Technology

NEURAL NETWORK-BASED ECHOCARDIOGRAM VIDEO CLASSIFICATION BY INCORPORATING DYNAMIC INFORMATION AND ATTENTION MECHANISM

YE ZI

Doctor of Philosophy

2022

NEURAL NETWORK-BASED ECHOCARDIOGRAM VIDEO CLASSIFICATION BY INCORPORATING DYNAMIC INFORMATION AND ATTENTION MECHANISM

YE ZI

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

Faculty of Information and Communication Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2022

DECLARATION

I declare that this thesis entitled “Neural Network-based Echocardiogram Video Classification by Incorporating Dynamic Information and Attention Mechanism” is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature :………………..……………

Name : YE ZI

Date : 28/09/2022

APPROVAL

I hereby declare that I have read this thesis and in my opinion, this thesis is sufficient in terms of scope and quality for the award of the degree of Doctor of Philosophy.

Signature : ………………………………

Supervisor Name : Dr. Yogan Jaya Kumar

Date : 28/09/2022

DEDICATION

This thesis is dedicated to the people who have supported me throughout my education.

Thanks for making me see this adventure through to the end.

ABSTRACT

Echocardiography,the use of ultrasound waves to investigate the heart's action, is the primary physiological test for cardiovascular disease diagnoses. The determination of the probe viewpoint forms an essential step in echocardiographic image analysis. Some of such views are identified as standard views due to the presentation and ease of the evaluations of their major cardiac structures. Finding valid cardiac views has traditionally been a laborious process and interpreted manually by the specialist, so there exists significant interest in providing an automated classification of echocardiograms in order to speed up the diagnosis process. However, the traditional machine learning methods require time-consuming and operator-dependent manual selection and annotation of features. Therefore, this study aims to simplify the diagnosis process by providing an automated classification of standard cardiac views based on deep learning technology. More importantly, our research considers and assesses some new neural network architectures driven by action recognition in video. For this aim, two classes of neural network architectures have been outlined: the CNN+BiLSTM model and the Spatiotemporal-BiLSTM model. It is finally verified that these methods aggregating dynamic information receive a more substantial classification effect. In addition, previous observations concluded that the most significant challenge lies in distinguishing among the various adjacent views. To this end, our study further aimed to adopt the attention mechanism for designing efficient neural networks. We proposed an ECHO-Attention architecture consisting of two parts. We first design an ECHO-ACTION block, which efficiently encodes Spatio-temporal features, channel-wise features, and motion features. Then, we insert this block into existing ResNet architectures, combined with a self-attention module to ensure its echocardiogram classification task-related focus, to form an effective ECHO-Attention network. All of these experiments are implemented on our privately collected dataset of 2693 videos acquired from 267 patients, which trained cardiologists have manually labeled. The evidence from this study showed that all the proposed methods yielded good results. The ECHO-Attention architecture provides the best classification performance (overall accuracy of 94.81%) on the entire video sample and achieved significant improvements on the classification of anatomically similar views (precision 88.65% and 81.70% for PSAX-AP and PSAX-MID on 30-frame clips, respectively).

KLASIFIKASI VIDEO EKOKARDIOGRAM BERASASKAN RANGKAIAN NEURAL DENGAN MENGGABUNGKAN MAKLUMAT DINAMIK DAN MEKANISME PERHATIAN

ABSTRAK

Ekokardiografi, yang menggunakan ultrabunyi untuk memeriksa fungsi jantung, adalah ujian fisiologi utama untuk diagnosis penyakit kardiovaskular. Penentuan sudut pandangan kuar adalah langkah penting dalam analisis imej ekokardiografi. Beberapa pandangan ini dikenal pasti sebagai pandangan piawai kerana tahap kemudahan penilaian dan pandangan struktur jantung utama. Secara tradisinya, mencari pandangan jantung secara berkesan menjadi satu proses yang sukar dan perlu ditafsirkan secara manual oleh pakar perubatan. Maka terdapat minat yang besar untuk menyediakan klasifikasi automatik ekokardiogram untuk mempercepatkan proses diagnostik. Walau bagaimanapun, kaedah pembelajaran mesin tradisional memakan masa dan bergantung pada pengendali untuk memilih dan menganotasi ciri secara manual. Oleh itu, kajian ini bertujuan untuk memudahkan proses diagnostik dengan menyediakan klasifikasi automatik pandangan jantung piawai berdasarkan teknik pembelajaran mendalam. Lebih penting lagi, kajian kami mempertimbangkan dan menilai beberapa seni bina rangkaian neural baharu yang didorong oleh pengecaman tindakan dalam video. Untuk tujuan ini, dua kelas seni bina rangkaian neural telah digariskan: model CNN+BiLSTM dan model Spatiotemporal-BiLSTM. Akhirnya, didapati kesan pengelasan yang diperolehi oleh kaedah mengumpulkan maklumat dinamik ini adalah lebih ketara. Tambahan pula, pendapat terdahulu menyimpulkan bahawa cabaran terbesar adalah dalam membezakan antara pelbagai sudut pandangan yang bersebelahan. Untuk tujuan ini, kajian kami selanjutnya menggunakan mekanisme perhatian untuk mereka bentuk rangkaian neural yang cekap. Kami mencadangkan seni bina ECHO-Attention yang terdiri daripada dua bahagian. Kami mula-mula mereka bentuk model ECHO-ACTION yang dapat mengekod ciri ruang-masa, ciri arah saluran dan ciri gerakan dengan cekap. Kemudian kami memasukkan blok ini ke dalam seni bina ResNet sedia ada, menggabungkan model perhatian kendiri untuk memastikan klasifikasi ekokardiografinya tugasan berfokus, bagi membentuk rangkaian ECHO-Attention yang cekap. Semua eksperimen ini dilakukan pada set data yang dikumpul secara peribadi dan mengandungi 2693 video daripada 267 pesakit yang dilabel secara manual oleh pakar kardiologi terlatih. Bukti daripada kajian ini menunjukkan bahawa semua kaedah yang dicadangkan mencapai hasil yang baik. Seni bina ECHO-Attention memberikan prestasi klasifikasi terbaik (ketepatan keseluruhan 94.81%) pada keseluruhan sampel video dan mencapai peningkatan ketara pada klasifikasi pandangan yang serupa secara anatomi (ketepatan 88.65% dan 81.70% untuk PSAX-AP dan PSAX-MID pada 30 bingkai klip, masing-masing).

ACKNOWLEDGEMENTS

Countless people supported my effort on this thesis. Thank you to my supervisors, Dr. Yogan Jaya Kumar and Prof. Ts. Dr. Goh Ong Sing, for your patience, guidance, and support. They provided invaluable feedback on my analysis and framing, at times responding to emails late at night and early in the morning. I have significantly benefited from your wealth of knowledge and meticulous editing.

Several other professors gave helpful advice as I wrote. Thank you to Fengyan Song, who generously took time out of his schedules to participate in my research and make this project possible. Thank you to Xianda Ni, for the professional assistance you have provided. Your encouraging words and thoughtful, detailed feedback have been very important to me. I would also like to acknowledge Pantai Hospital Ayer Keroh for valuable support regarding data resources.

Most importantly, I am grateful for my family's unconditional, unequivocal, and loving support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

2D	-	Two-Dimensional
A2C	-	Apical Two-Chamber View
A3C	-	Apical Three-Chamber View
A4C	-	Apical Four-Chamber View
A5C	-	Apical Five-Chamber View
aCMQ	-	Automated Cardiac Motion Quantification
ACTION	-	spAtio-temporal, Channel and moTion excitatION
ADAM	-	Adaptive Moment Estimation
AI	-	Artificial Intelligence
AVA	-	Aortic Valve Opening Area
BiLSTM	-	Bidirectional Long Short-Term Memory
BN	-	Batch Normalization
BoW	-	Bag-Of-Words
BP	-	Background Propagation
CAD	-	Computer-Aided Diagnosis
CE	-	Channel Excitation
CM	-	Channel Module
CNN	-	Convolutional Neural Network
CPU	-	Central Processing Unit
CT	-	Computed Tomography
CVD	-	Cardiovascular Diseases

DICOM	-	Digital Imaging And Communications In Medicine
DL	-	Deep Learning
ECG	-	Electrocardiogram
EF	-	Ejection Fraction
FC	-	Fully Connected Layer
FN	-	False Negative
FoV	-	Field Of View
FP	-	False Positive
GLM	-	Generalized Linear Model
GPU	-	Graphics Processing Unit
IRB	-	Institutional Review Board
LSTM	-	Long Short-Term Memory
MB-SGD	-	Mini-Batch Stochastic Gradient Descent
ME	-	Motion Excitation
MINIST	-	Modified National Institute Of Standards And Technology
ML-DNN	-	Multi-Loss Deep Neural Network
MLP	-	Multilayer Perceptron
MM	-	Motion Module
NIN	-	Network In Network
NLP	-	Natural Language Processing
PLAX	-	Parasternal Long-Axis View Of The Left Ventricle
PSAX	-	Parasternal Short-Axis View
PSAX-AP	-	Parasternal Short-Axis View Of The Left Ventricle At The Apical Level
PSAX-AV	-	Parasternal Short-Axis View Of The Aorta
PSAX-MID	-	Parasternal Short-Axis View Of The Left Ventricle At Papillary Muscle Level
PSAX-MV	-	Parasternal Short-Axis View Of The Left Ventricle At The Mitral

		Valve Level
RNN	-	Recurrent Neural Network
SC	-	Subcostal
SSN	-	Suprasternal
STE	-	Spatio-Temporal Excitation
STM	-	Spatio-Temporal Module
STN	-	Spatial Transformer Network
SVHN	-	Street View House Numbers
SWWAE	-	Stacked What-Where Auto-Encoders
TN	-	True Negative
TP	-	True Positive
US	-	Ultrasound
VGG	-	Visual Geometry Group
VTN	-	Video Transformer Network