A Confidence Framework for Heart Rate Estimation in Video Magnification

by

Diane Elhajjar

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Applied Science

in

Electrical and Computer Engineering

Carleton University
Ottawa, Ontario

© 2023, Diane Elhajjar
Copyright

Parts of this thesis were published in “Assessing Confidence in Video Magnification Heart Rate Measurement using Multiple ROIs” [1] and “A Confidence Framework through Temporal Averaging for Heart Rate Estimation in Video Magnification” [2]. The above-stated references were presented at and are in conference proceedings for conferences by the Institute of Electrical and Electronics Engineers (IEEE). IEEE allows the reuse of published articles in a thesis or dissertation without requesting permission (see Declaration message from IEEE below). The material used from the above conference proceedings [1]–[2] will be both referenced in the following chapters as © 2023 [1] for [1] and © 2023 [2] for [2]. I am the lead author on both conference proceedings stated above.

Declaration message from IEEE:

In reference to IEEE copyrighted material which is used with permission in this thesis, IEEE does not endorse any of Carleton University’s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.
Abstract

In-home health monitoring solutions are becoming increasingly necessary as a result of both the recent pandemic and the growing search for solutions to support independent living for aging adults. Video magnification (VM) is a contactless and remote method that measures heart rate (HR), which is a vital parameter and an indicator of the overall health and well-being of individuals. Body motion, illumination, and skin color can affect VM performance, and this thesis proposes a methodology to understand the confidence in VM heart rate assessments. The thesis proposes both spatial methods, combining VM assessment from different skin regions in time and frequency domains, and temporal methods, combining VM assessments from adjacent time windows, to improve VM HR estimation and provide confidence assessments. The thesis then proposes the use of Machine Learning models to assess the overall confidence in the VM assessed HR to indicate whether a HR is likely correct or incorrect.
Acknowledgements

First and foremost, I would like to thank my research supervisors, Dr. Rafik Goubran and Dr. Bruce Wallace, for providing invaluable support throughout my research journey and the full thesis as a whole. I am truly grateful for their effort, time, and commitment in maintaining a wonderful, healthy, and professional environment which has been key to my development and performance. Both supervisors have been crucial to my success as I have learnt a lot from their experience, as they are well versed in the research field throughout my thesis. I am thankful for Dr. Goubran for his kindness, generosity, and sincere follow ups whether intellectually, academically, or financially. Additionally, I would like to sincerely thank Dr. Wallace for being a great mentor and supervisor since day one. Dr. Wallace has been tremendously helpful and supportive from the continuous advice and resources that he has provided to his willingness to respond to any of my inquiries. With the support from both, I was able to provide meaningful contribution to the realm of the research project and to the team.

Furthermore, I want to thank the Faculty of Engineering and Design at Carleton University for providing me with an immense scholarship that greatly impacted my journey both academically and personally. I am extremely honored that I was selected as a recipient of the research assistance award that was supported by the National Research Council (NRC) to fully fund the tuition of my master’s program, which has provided me with the financial stability to not only pursue the master’s program but also perform well through its significant impact on my mental and emotional wellbeing. Additionally, I am honored that I was chosen as the recipient of the Anita Garbarino Girard Fund for Fall
2022m which has made a positive impact on my accomplishments and performance, especially financially supporting me during inevitable hard times.

On a separate note, I would like to thank Dr. Andrew Law at the NRC for his assistance, guidance, and continuous feedback on the technical and research aspects which helped me present my work at its best. Furthermore, I want to acknowledge the support of both Dr. Frank Knoefel and Dr. Neil Thomas at the Bruyère Research Institute for their diversified knowledge and experience in the fields of medical practice and research. I would like to thank as well the research laboratory technologist, Mr. Julien Lariviere-Chartier for his technical support, time, and effort that he provided to me during the preliminary stages of my research. Finally, I would like to thank my family for their continuous support and motivation that they have provided along my research journey.
Table of Contents

Abstract ........................................................................................................................................ iii
Acknowledgements .................................................................................................................... iv
Table of Contents ....................................................................................................................... vi
List of Tables .............................................................................................................................. vi
List of Figures ............................................................................................................................. xi
List of Appendices ....................................................................................................................... xvi
List of Acronyms ........................................................................................................................ xvii

Chapter 1: Introduction .................................................................................................................. 19
  1.1 Background ............................................................................................................................ 19
  1.2 Motivation ............................................................................................................................. 20
  1.3 Objectives .............................................................................................................................. 22
  1.4 Summary of Main Contributions ......................................................................................... 23
  1.5 Summary of Collaborations ............................................................................................... 24
  1.6 Thesis Structure .................................................................................................................... 25

Chapter 2: Background and Literature Review ............................................................................. 26
  2.1 Primary HR Assessment through Contact Methods ............................................................ 26
  2.2 Contactless and Remote Heart Rate Assessment ............................................................... 28
    2.2.1 Eulerian Video Magnification (EVM) and Remote Photoplethysmography (r-PPG) for
         Contactless and Remote HR Assessment .................................................................... 28
    2.2.2 Existing Review Articles on Remote Heart Rate Assessment .................................. 30
  2.3 Foundation Methods for Optical HR Assessment .............................................................. 32
  2.4 Machine Learning ............................................................................................................... 34
  2.5 Challenges with Video Magnification ................................................................................. 35
  2.6 Confidence in HR Measurement ......................................................................................... 36
2.7 Spatial Domain within Video Magnification ................................................................. 37
2.8 Temporal Domain within Video Magnification ........................................................... 38

Chapter 3: Experimental Set-up .................................................................................. 39
3.1 Dataset ....................................................................................................................... 39
3.2 VM implementation method ...................................................................................... 42
  3.2.1 Time Windowing Method .................................................................................... 52
3.3 Summary .................................................................................................................. 54

Chapter 4: Confidence Framework through Spatial Domain Methods ............... 56
4.1 Introduction ............................................................................................................... 56
4.2 Methods ...................................................................................................................... 56
  4.2.1 Base VM Spatially Across Multiple ROIs ............................................................ 56
  4.2.2 Spatial Frequency Domain Averaging (SFDA) Method ......................................... 58
  4.2.3 Spatial Time Domain Averaging (STDA) Method ............................................... 59
4.3 Results ....................................................................................................................... 60
  4.3.1 Results for Base VM Across Multiple ROIs .......................................................... 61
    4.3.1.1 Group A: Cases with Good Performance ....................................................... 61
    4.3.1.2 Group B: Cases with Weak Performance ....................................................... 63
    4.3.1.3 Group C: Case with Poor Performance ....................................................... 67
  4.3.2 Results for the SFDA Method .............................................................................. 71
  4.3.3 Results for STDA Method ................................................................................. 83
  4.3.4 Summary Results for All 19 Subjects ................................................................. 90
4.4 Discussion .................................................................................................................. 92

Chapter 5: A Confidence Framework through Temporal Domain Methods ........ 95
5.1 Introduction ............................................................................................................... 95
5.2 Methods ..................................................................................................................... 95
5.2.1 Application of Base VM Temporally Across Multiple Time Windows .......... 95
5.2.2 Temporal FFT Averaging (TFA) Method ........................................... 97
5.3 Results ................................................................................................. 99
5.3.1 Results for Base VM Method Temporally Across Multiple Time Windows .... 99
  5.3.1.1 Group A: Cases with Good Performance ..................................... 100
  5.3.1.2 Group B: Cases with Weak Performance .................................... 101
  5.3.1.3 Group C: Cases with Poor Performance ...................................... 102
5.3.2 Results for Temporal FFT Averaging (TFA) Method ............................ 103
  5.3.2.1 Group A: Cases with Good Performance ..................................... 104
  5.3.2.2 Group B: Cases with Weak Performance .................................... 107
  5.3.2.3 Group C: Cases with Poor Performance ...................................... 110
  5.3.2.4 Summary of Combined Results .................................................. 113
5.4 Discussion ............................................................................................ 114

Chapter 6: Confidence Assessment of VM HR through Machine Learning ........ 116
6.1 Introduction ........................................................................................... 116
6.2 Methods ................................................................................................ 117
  6.2.1 Feature Assignment for Spatial Time Domain Averaging (STDA) Method .... 118
  6.2.2 Feature Assignment for Temporal FFT Averaging (TFA) Methods ............... 119
  6.2.3 Features Assignment for Base VM ....................................................... 121
  6.2.4 Application of Machine Learning and Features Selection ....................... 122
    6.2.4.1 Feature Selection ........................................................................ 122
    6.2.4.2 Importing Feature Dataset into Training Models .......................... 124
6.3 Results .................................................................................................... 124
  6.3.1 Results for Base VM Method .............................................................. 125
  6.3.2 Results for STDA Method ................................................................. 125
  6.3.3 Results for TFA Method .................................................................... 127
6.3.4 Results from Combining STDA and TFA Features Dataset ........................................ 127

6.4 Discussion .................................................................................................................. 133

Chapter 7: Conclusion ........................................................................................................ 135

Appendices ......................................................................................................................... 138

Appendix A ......................................................................................................................... 138

A.1 Mapping of NRC Subjects ....................................................................................... 138

Appendix B ......................................................................................................................... 140

B.1 Code Stack for ROI Selection and Acquisition ....................................................... 140

References .......................................................................................................................... 141
List of Tables

Table 4.1. Largest peak indicated HR and FFT_Ratio for SFDA and STDA methods compared to the FFT_Ratio for R1 alone and to the true subject HR via PPG as well as indicating improved performance in HR prediction (in green) and poor results (in red). 92

Table 5.1. Largest and second largest peak HR with normalized magnitude, and FFT_Ratio for method 1 and TFA for HR within ±1 BPM (case 1), ±2 BPM (case 2), ±3 BPM (case 3) for subject 2, compared to PPG HR. © 2023 [2] 107

Table 5.2. Largest and second largest peak indicated HR with normalized magnitude, and FFT_Ratio for method 1 and TFA for HR within ±1 BPM (case 1), ±2 BPM (case 2), ±3 BPM (case 3) for subject 1, and compared to PPG HR. © 2023 [2] 109

Table 5.3. Largest and second largest peak indicated HR with normalized magnitude, and FFT_Ratio for method 1 and TFA for HR within ±1 BPM (case 1), within ±2 BPM (case 2), within ±3 BPM (case 3) for subject 16, compared to PPG HR. © 2023 [2] 112

Table 5.4. Minimum, maximum, average, and standard deviation for the FFT_Ratio across windows 2 through 6 for methods 1 and the 3 cases for TFA for the choice of ± m BPM for all 19 subjects among groups A, B, and C, and the percentage of windows for HR within ±2BPM. © 2023 114

Table 6.1. Feature name and description for STDA method 119

Table 6.2. Feature name and description for TFA method 120

Table 6.3. Feature name and description of confidence measure for base VM method 121

Table 6.4 Accuracy Score for training models for base VM method 125

Table 6.5. Accuracy Score for training models for STDA method 126

Table 6.6. Accuracy Score for training models on selected feature sets for all 3 cases of TFA method 127

Table 6.7. Accuracy Score for training models for features sets for combined STDA and TFA 128
List of Figures

Fig. 3.1 Overview of the experimental setup for data collection by NRC under one specific test condition; Video captures using RGB camera with subjects seated, no mask, LED light source, as well as having PPG acquisition alongside [1], [27].........................42

Fig. 3.2. Experimentation and test conditions set up for video and PPG acquisition to test video processing for video magnification algorithm.........................................................43

Fig. 3.3. Block diagram for the framework of the base VM method..................................45

Fig. 3.4. Structuring all the imported video frames of video into a 2-D matrix array...........45

Fig. 3.5. Illustration showing the 3 Red-Green-Blue channels for the image frame in......46

Fig. 3.6. Illustration of ROI selection on a single inner frame using - marker points on the forehead skin region. ........................................................................................................47

Fig. 3.7. Generic flow diagram for computing average pixel for a selected single ROI...47

Fig. 3.8. Illustration of time-domain processing of the ROI stack involving band pass filtering to preserve the HR signals .................................................................48

Fig. 3.9. Low pass filter using Park's MacLellan FIR Filter with corner [3, 3.5] Hz........49

Fig. 3.10. High pass filter using Park's MacLellan FIR Filter with corner [0.45, 0.65] Hz. ..........................................................................................................................49

Fig. 3.11. Mean Pixel value across time for a ~90 sec. video before filtering (top) and after bandpass filtering with a low pass filter (middle) followed by a high pass filter (bottom)........................................................................................................50

Fig. 3.12. Transformation from time domain to frequency domain using Fast Fourier Transform (FFT) for Heart Rate (HR) measurement .......................................................51

Fig. 3.13. FFT result plot indicating a largest peak at ~52 BPM........................................51

Fig. 3.14. Generic block diagram for VM on RC dataset (Video and PPG data) obtained from the 19 NRC subjects.......................................................................................52

Fig. 3.15. VM application on multiple time windows through a hamming window convolved with a 20 sec. time segments with 50% overlap.................................................54

Fig. 4.1. Generic algorithmic model for the application of base VM spatially across multiple ROIs for spatial redundancy..............................................................................58
Fig. 4.2. Block diagram for measuring heart rate through VM using frequency domain averaging (SFDA) method for a 20 sec. video. © 2023 [1]

Fig. 4.3. Block diagram for heart rate measurement using time domain averaging (STDA) method for a 20 sec. video. © 2023 [1]

Fig. 4.4. FFT plot for VM on ROI R1 (forehead) for a 20 sec. of video (window 1) for subject 3 with PPG HR of ~52 BPM.

Fig. 4.5. FFT plot for VM on ROI R2 (left cheek) for a 20 sec. of video (window 1) for subject 3 with PPG HR of ~52 BPM.

Fig. 4.6. FFT plot for VM on ROI R3 (right cheek) for a 20 sec. of video (window 1) for subject 3 with PPG HR of ~52 BPM.

Fig. 4.7. FFT plot for VM on ROI R1 (forehead) for a 20 sec. of video (window 4) for subject 7 with PPG HR of ~60 BPM.

Fig. 4.8. FFT plot for VM on ROI R2 (left cheek) for a 20 sec. of video (window 4) for subject 7 with PPG HR of ~60 BPM.

Fig. 4.9. FFT plot for VM on ROI R3 (nose) for a 20 sec. of video (window 4) for subject 7 with PPG HR of ~60 BPM.

Fig. 4.10. FFT plot for VM on ROI R1 (forehead) for a 20 sec. of video (window 5) for subject 11 with PPG HR of ~86 BPM.

Fig. 4.11. FFT plot for VM on ROI R2 (left cheek) for a 20 sec. of video (window 5) for subject 11 with PPG HR of ~86 BPM.

Fig. 4.12. FFT plot for VM on ROI R3 (nose) for a 20 sec. of video (window 5) for subject 11 with PPG HR of ~86 BPM.

Fig. 4.13. FFT plot for VM on ROI R1 (forehead) for a 20 sec. of video (window 3) for subject 16 with PPG HR of ~68 BPM.

Fig. 4.14. FFT plot for VM on ROI R2 (left cheek) for a 20 sec. of video (window 3) for subject 16 with PPG HR of ~68 BPM.

Fig. 4.15. FFT plot for VM on ROI R3 (right cheek) for a 20 sec. of video (window 3) for subject 16 with PPG HR of ~68 BPM.

Fig. 4.16. FFT plot for VM on ROI R4 (nose) for a 20 sec. of video (window 3) for subject 16 with PPG HR of ~68 BPM.

Fig. 4.17. The Average of all 3 FFT for the same example subject in Fig. 4.7, 4.8, and 4.9.
Fig. 4.18. FFT results of VM for subject 17 for 20 sec. of video (time window 5) for ROI R1, R2, R3, and R4. The PPG HR is 47.8 BPM. ..........................................................72

Fig. 4.19. FFT result plot of VM for the same example subject in Fig. 4.18 using SFDA method. ........................................................................................................73

Fig. 4.20. FFT results of base VM for subject 1 for ROI R1, R2, R3, and R4 for 20 sec. of video (time window 5). The PPG HR is 66.5 BPM. ..........................................................74

Fig. 4.21. FFT result plot of VM for the same example subject in Fig. 4.20 using SFDA method. ........................................................................................................75

Fig. 4.22. FFT results of VM for subject 19 for 20 sec. of video (time window 2) for ROI R1, R2, R3, and R4. The PPG HR is 84.1 BPM. ..........................................................76

Fig. 4.23. FFT result plot of VM for the same example subject in Fig. 4.22 using SFDA method. ........................................................................................................77

Fig. 4.24. FFT results of VM for subject 11 for 20 sec. of video (time window 4) for ROI R1, R2, R3, and R4. The PPG HR is 82.9 BPM. ..........................................................78

Fig. 4.25. FFT result plot of VM for the same example subject in Fig. 4.24 using SFDA method. ........................................................................................................79

Fig. 4.26. FFT results of VM for subject 12 for 20 sec. of video (time window 4) for ROI R1, R2, R3, and R4. The PPG HR is 65.3 BPM. ..........................................................80

Fig. 4.27. FFT result plot of SFDA method for the same example subject in Fig. 4.26. ..........................................................81

Fig. 4.28. FFT results of VM for subject 18 for 20 sec. of video (time window 4) for ROI R1, R2, R3, and R4. The PPG HR is 91.7 BPM. ..........................................................82

Fig. 4.29. FFT result for SFDA method for the same example subject in Fig. 4.28. ..........................................................83

Fig. 4.30. FFT result plot of VM for the same example subject in Fig. 4.18 using STDA method. ........................................................................................................84

Fig. 4.31. FFT result plot of VM for the same example subject in Fig. 4.20 using STDA method. ........................................................................................................85

Fig. 4.32. FFT result plot of VM for the same example subject in Fig. 4.22 using STDA method. ........................................................................................................86

Fig. 4.33. FFT result plot for the same example subject in Fig. 4.24 using STDA method. ........................................................................................................87
Fig. 4.34. FFT result plot for the same example subject in Fig. 4.26 using STDA method.

Fig. 4.35. FFT result plot for the same example subject in Fig. 4.28 for STDA method.

Fig. 5.1. Block Diagram of VM Model in multiple time windows for measuring FFT_Ratio. © 2023 [2]

Fig. 5.2. Block Diagram for computing the average FFT_Ratio over all 7 windows. © 2023 [2]

Fig. 5.3. Block Diagram for HR Measurement using TFA method. © 2023 [2]

Fig. 5.4. VM results on a single ROI (forehead) for subject 10 for time windows 3, 4, and 5 for with respective PPG HR of ~48 BPM.

Fig. 5.5. VM results on a single ROI (forehead) for subject 4 for time windows 1, 2, and 3 with respective PPG HR of ~81 BPM.

Fig. 5.6. VM results on a single ROI (forehead) for subject 13 for time windows 4, 5, and 6 with respective PPG HR of ~68 BPM.

Fig. 5.7. Example Video Magnification results for base VM method (method 1) for subject 2 (Group A) for the three 20-second time windows of video using the region of interest shown in Fig. 5.1. The true PPG HR is 64.3 BPM for window 2, 63.2 BPM for window 3, and 64.8 BPM for window 4. © 2023 [2]

Fig. 5.8. Example Video Magnification results for TFA case 2 for subject 2 (Group A) for three 20-second time windows of video using the same region of interest shown in Fig. 5.1. The true PPG HR is 64.3 BPM for window 2, 63.2 BPM for window 3, and 64.8 BPM for window 4. © 2023 [2]

Fig. 5.9. Example Video Magnification results for method 1 for subject 1 (Group B) for three 20-second time windows of video using the region of interest shown in Fig. 5.1. The true PPG HR for window 3, 4, and 5 is ~65.0, ~67.0, and ~66.0 BPM. © 2023 [2]

Fig. 5.10. Example Video Magnification results for TFA case 2 for subject 1 (Group B) for three 20-second time windows of video using the region of interest shown in Fig. 5.1. The true PPG HR for window 3, 4, and 5 is ~65.0, ~67.0, and ~66.0 BPM. © 2023 [2]

Fig. 5.11. Example Video Magnification results for method 1 for subject 16 (Group C) for three 20-second time windows of video using the region of interest shown in Fig. 5.1. The true PPG HR is 71.9 BPM for window 4, 77.4 BPM for window 5, and 73.1 BPM for window 6.
Fig. 5.12. Example Video Magnification results for TFA case 2 for subject 16 (Group C) for three 20-second time windows of video using the region of interest shown in Fig. 5.1. The true PPG HR is 71.9 BPM for window 4, 77.4 BPM for window 5, and 73.1 BPM for window 6. © 2023 [2]..................................................................................................................112

Fig. 6.1. Machine learning block diagram for HR classification. ..............................118

Fig. 6.2. Overview of the prepared table for machine learning with all the features in Table 6.1 for STDA method. .................................................................................................119

Fig. 6.3. Data Set for all the features in Table 6.2 for the 3 cases of TFA methods. ......121

Fig. 6.4. Overview of the feature data set table from Table 6.3 for the base VM method. .................................................................................................................................122

Fig. 6.5. Assigning all features from STDA, TFA, and base VM across distinct pools based on the confidence measure. .....................................................................................123

Fig. 6.6. Illustration of a summary table with all the dataset of the STDA and TFA methods. ..........................................................................................................................124

Fig. 6.7. Confusion Matrix for observatory performance of the Fine Tree model on STDA feature dataset..................................................................................................................126

Fig. 6.8. Confusion matrix for Linear SVM with 90% accuracy score for STDA and TFA feature set ...................................................................................................................129

Fig. 6.9. Confusion matrix with 92.6% accuracy score using SVM kernel for STDA and TFA feature set ..............................................................................................................130

Fig. 6.10. Confusion matrix for 90% accuracy score using Fine Tree model for STDA and TFA feature set .................................................................................................130

Fig. 6.11. Confusion matrix for 78.9% accuracy test score using Linear SVM model for STDA and TFA feature set. .....................................................................................131

Fig. 6.12. Confusion matrix for 82.1% accuracy test score using SVM Kernel model for STDA and TFA feature set. .....................................................................................131

Fig. 6.13. Confusion matrix for 89.5% accuracy test score using SVM Kernel model for STDA and TFA feature set. .....................................................................................132

Fig. 6.14. Confusion matrix for 82.1% accuracy test score using Fine Tree model for STDA and TFA feature set. .....................................................................................132
List of Appendices

Appendix A ................................................................................................................................. 139
    A.1 Mapping of NRC Subjects .................................................................................................. 139
Appendix B ..................................................................................................................................... 141
    B.1 Code Stack for ROI Selection and Acquisition ................................................................. 141
List of Acronyms

\textit{Avg}_{\text{FFT Ratio}}: Average across all FFT Ratio

\textbf{BPM}: Beats Per Minute

\textbf{EVM}: Eulerian Video Magnification

\textbf{FFT}: Fast Fourier Transform

\textbf{FFT}_{w}: FFT for current time window

\textbf{FFT}_{w,\text{new}}: FFT with new results for current time window

\textbf{FFT}_{w-1}: FFT for preceding time window

\textbf{FFT}_{w+1}: FFT for subsequent time window

\textbf{FFT Ratio}: Ratio between \text{NM}_{LP} and \text{NM}_{SLP}

\textbf{HR}: Heart Rate

\textbf{HR}_{LP}: Largest Peak Indicated Heart Rate

\textbf{HR}_{SLP}: Second Largest Peak Indicated Heart Rate

\textbf{LP}: Largest Peak

\textbf{NM}_{LP}: Largest Peak Indicated Normalized Magnitude

\textbf{NM}_{SLP}: Second Largest Peak Indicated Heart Rate

\textbf{NRC}: National Research Council

\textbf{PPG}: Photoplethysmography

\textbf{RGB}: Red/Green/Blue

\textbf{ROI}: Region of Interest

\textbf{r-PPG}: Remote Photoplethysmography

\textbf{SFDA}: Spatial Frequency Domain Averaging

\textbf{SLP}: Second Largest Peak
**STDA**: Spatial Time Domain Averaging

**TFA**: Temporal FFT Averaging

**VM**: Video Magnification
Chapter 1: Introduction

This chapter includes a background section on Heart Rate (HR) as an essential vital sign, a motivation section that introduces the base Video Magnification (VM) method for remote and contactless HR assessment, and an objective section that highlights the two main proposed methods that are implemented to create the confidence framework. From there it provides a summary of contributions highlighting major research work through distinct publications, followed by a summary of collaboration section that summarizes all the collaborative work that was performed by all external and internal members of the research team.

1.1 Background

HR is an essential vital sign that is primarily used for assessing the health of individuals, and it is considered as an essential element in the diagnosis and treatment of cardiovascular diseases [3]–[9]. HR is measured as the total number of heartbeats occurring in one minute which can, through any irregularities, provide insights on the physiological functioning and the wellbeing of the heart [6], [9], [10]–[12]. For most individuals, HR typically ranges between 30 to 180 beats per minute (BPM), but it can reach into the low 200s. HR can indicate cardiovascular complications that can be associated with different types of physiological disorders such as hypertension, allowing it to be a crucial component in health monitoring systems and primary care [13], [14]. For example, HR can directly indicate cardiac disorders such as arrhythmia by the irregularities of HR caused by the improper coordination of the cardiac cycle activity controlled by sinus node electrical signals [6], [15]–[18]. In addition to arrhythmia, HR can be used to
indicate serious medical conditions which can be linked to the cardio-vascular fitness, and requires serious attention as it can cause potentially fatal heart attacks [12].

The gold standard for heart rate assessment is electrocardiography (ECG), and, while there are other clinical HR measurement techniques (e.g. pulse oximetry), they all require direct application of a sensor to the skin [5], [8], [19]. The complexity in operating and accessing ECG devices requires visits to health care centres as well as intervention from medical and health practitioners [20]. Additionally, pulse oximeters require careful use to ensure proper and accurate HR measurement, as they can be highly prone to motion effects and it can take some time to stabilize the HR measure before they provide an accurate HR [21]. In addition to the limitations, the recent COVID-19 pandemic and the increased focus towards senior independent living have increased the need to augment these traditional measuring devices using advanced technology [3], [8]. Therefore, research focused on advancing technological solutions that can be incorporated into remote methods such as digital health monitoring systems is valuable [2], [22], [23], [24].

1.2 Motivation

VM is a contactless and remote method that can measure HR by magnifying subtle variations captured on video associated with blood flow [25], [26]. There are changes in skin color are caused by the presence (during systole) and absence (during asystole) of blood in the capillaries of the skin [27]. These changes are minuscule and cannot be detected by the naked eye because they are close to the noise levels for the video camera [25], [27], [28]. However, the HR associated with the color changes is correlated across the pixels within a skin region, and therefore, VM algorithms combine the pixels within
a skin region and process the sequence of video frames to magnify the color variation and allow for the measurement the HR [25], [29].

VM has numerous advantages as it is safe and quick, applicable for neonatal health monitoring purposes, and convenient to those living far from healthcare centres. Moreover, with its remote capabilities, VM can play an important role in providing support to remote communities such as rural areas, which often have aging populations and tend to have limited access to medical support [4], [23], [26], [27].

On the other hand, VM has a number of limitations that can adversely affect its performance, leading to improper HR measurement. As VM is a video-based method, it can be highly sensitive to various factors and research has indicated that body motion, illumination conditions, and skin color can prevent proper HR assessment, and these issues result in interferences that tend to appear at the same noise floor of the camera that captures the video [1], [2], [18], [27], [28], [30]–[34]. These limitations can obscure the true variations of the color changes that occur in video, leading to inaccurate HR measures [1], [23], [28], [29].

VM could be a component in remote HR assessment solutions which have lead many researchers to explore different types of disciplinary approaches to reduce and mitigate the effects of the limitations [23], [25], [27], [28], [30], [31]. However, alongside addressing solutions to those limitations, there was very minimal focus in building levels of confidence around the correct HR. VM is a 2-staged algorithm involving time and frequency domains to allow a measure of HR from a series of video processing. Many researchers explored areas in VM through the implementation of advanced methods [23], [28], [35]. However, the solutions were found to be both non trivial and not robust for
all cases. Furthermore, the main focus of previous work was on specific limitations rather than in general.

Indeed, there is a high need to address limitations in VM by exploring methods that will help identify the correct HR in the presence of other conflicting VM measures. Minimal research looked into reliability measurements in VM, as most of it had limited focus in which all the implemented solutions and methods were related to a specific limitation. Even with improved methods, there was no evidence in regards to the reliability of the HR measures obtained. Therefore, it is highly necessary to build levels of confidence around the measured HR regardless if it is correct or incorrect [1], [2]. It is not enough to reduce limitations as it is essential to be able to accurately indicate whether a measured HR is correct or not [1], [2], [31]. The reliability would allow researchers to explore redundancy measurements that would help identify and explore behavior of VM, guiding their understanding of the limitations and helping them reduce their adverse effects [1], [2].

1.3 Objectives

This thesis proposes a confidence framework for the estimation of HR by improving the identification of the correct HR from various VM results, making it possible to ignore and disregard incorrect and inaccurate ones. The research work explores spatial and temporal domain methods in VM in which the methods explore the spatial and temporal redundancies and evaluate the responses of VM across those redundancies. The spatial domain methods combine multiple regions of interest (ROIs) and the temporal domain methods combine the VM results from multiple and overlapping time intervals. The framework derives confidence measures leading to the proper HR estimation. Another
challenge with current VM methods is understanding the confidence in any VM-predicted HR. This thesis explores methods in the temporal and spatial domains to assess the confidence that a given HR prediction is likely to be correct or likely to be incorrect.

1.4 Summary of Main Contributions

The following briefly lists the 3 main contributions:

1- Spatial methods to combine multiple ROIs and to create initial confidence measures for the overall confidence framework to improve the HR estimation. A conference paper for this contribution is fully accepted and presented at the IEEE I2MTC conference for 2023.


2- Temporal methods to combine adjacent time intervals within a video and to build and model confidence measures for the overall confidence framework to improve the HR estimation. A conference paper for this contribution is accepted for presentation at the IEEE SAS conference, 2023.

3- Confidence measure combination from the spatial and temporal confidence framework to assess the confidence in the VM results using Machine Learning.

1.5 Summary of Collaborations

This section summarizes the collaborations that were essential for the successful completion of this thesis. My research supervisors, Dr. Rafik Goubran and Dr. Bruce Wallace, provided continuous advice as per their diversified expertise and knowledge on the research field. Dr. Wallace was the main technical lead for all the research work and implementations which guided and helped me in bridging my ideas and work into a whole thesis work. Moreover, they have constantly provided all the necessary theory and knowledge on the research topics and principles as well as guidance and support on the practical implementation aspect through the evaluation of the created algorithms and programs. Dr. Andrew Law from NRC has contributed on several aspects, from managing and performing the data collection to providing collaborative advice on various thesis topics through feedback on conference publications or by internal communication through during team meetings. Dr. Law played an essential role in helping me properly bridge ideas between the different topics of this thesis. Dr. Frank Knoefel contributed to the thesis by providing medical advice to ensure proper medical assumptions align with the methods and research goal. Finally, Mr. Julien Lariviere-Chartier provided laboratory support contribution by both preparing, the whole dataset by organizing all the raw data that was collected from NRC as well as offering me the guidance and support for proper accessibility for the dataset.
1.6 Thesis Structure

The thesis includes 7 chapters following the introduction (Chapter 1). Chapter 2 covers the background and literature review that focuses on the early research work on VM foundations and exploration on many disciplinary aspects to help improve its performance. Next, Chapter 3 explains the experimental setup tools and procedures that involved the dataset collection, preparation, and management that were involved as part of the implementation and testing of all the proposed methods of the thesis. Chapter 4 focuses on the implementations of the confidence framework in the spatial domain through spatial combination, and Chapter 5 focuses on the implementation of the confidence framework in the temporal domain through temporal combination. Chapter 6 introduces the Machine Learning framework to build a confidence framework through the combination of both spatial and temporal domain methods for an improved and a more precise confidence framework for the estimation of HR. Finally, a conclusion section summarizes the main contributions and lists the potential future areas of studies to be explored as part of the confidence assessment of HR in VM.
Chapter 2: Background and Literature Review

This chapter aims to present a review of the foundations of remote and contactless HR assessment methods. The review focuses on the early work in regard to HR measurement techniques to the establishment of optical HR assessment approach. It discusses the foundations of the VM methodologies, VM’s significance, and its limitations, and reviews the previous work that explored those limitations through various multi-disciplinary approaches.

2.1 Primary HR Assessment through Contact Methods

HR is a measure of the number of heartbeats that occur within a specific time interval, such as 60 seconds [15]. Cardiac activity is divided into diastole (heart refilling with blood) and systole (heart compression) periods which together make up the rhythm for normal functioning of the heart [3], [4], [8], [10], [35]. The heart is a muscle, and the pumping mechanism needs to be properly coordinated. For a proper cardiac cycle, the flow of blood into and through the heart chambers must be well coordinated through specific electrical impulses that guide the activity of the heart [38]. In the event of failure to maintain proper coordination, the heart cycle tends to be non-rhythmic, leading to irregular heartbeats which can warn for serious health conditions and physiological failures [6], [11], [16], [36].

An electrocardiogram (ECG) is the clinical gold standard method to assess the heart rate, measuring the bio-signals through direct application of electrodes on human skin [17], [19], [20], [39]. The bio-signals are acquired from a number of electrodes to obtain readings from various projections of the heart, leading to a HR measure. The ECG represents the signals as waveforms that are identified to properly measure the HR [40].
For a normal individual, the waveforms appear regular across the spectrum and any irregularities are indicative of cardiac failure to certain extent [5], [9], [10], [40], [41]. An ECG is able to detect the irregularities both directly by visualizing the waveforms, as well as through computational methods, making it a tool in the diagnosis and prognosis of diseases or health-related complications [15]. However, an ECG can also be affected by noise associated with contact motion when the subject moves [16], [39]. The ECG is highly prone to motion artifact, and it is shown to be non-robust to external noise signals which can interfere with the quality of the ECG signals, leading to inaccurate HR measurement [19], [39], [42], [43].

Other HR measurement technologies focus on different sensor modalities like optical-sensor photoplethysmography (PPG) [34], [44], [45]. Unlike an ECG, PPG acquires signals using optical sensors by transmitting light into a selected skin region where it interacts and is sent back to a photodetector [3], [20], [31], [41], [43]. PPG monitors the response and characteristics of the received light in order to measure and assess HR as well as oxygen saturation, and it relies on the basis of blood perfusion through the systole. PPG has been a reliable measurement tool for HR in clinical and healthcare settings for various medical purposes [45]. However, it does involve contact with the skin which makes it unfavorable in terms of safety and hygiene which are essential measures of proper health wellbeing [1], [2], [25], [45], [46]. PPG is highly accessible and more convenient for in-home applications in comparison to ECG. PPG are simple to use in comparison to ECG that involves the application of a complex sensory matrix which requires intervention of external healthcare units or support allowing it to
be less accessible. PPG does not require sticking electrodes directly to the skin which may have adverse effects such as causing skin irritations [47].

PPG has also shown limitations in regards to settling time in comparison to ECGs, and it has shown non-robustness in terms of body motion, occlusion of skin regions, and light interference [35], [44], [48]. With the advantages of both ECGs and PPG, they are considered accurate HR measurement techniques; however, various research work focused on incorporating advanced digital tools into health monitoring devices is necessary to improve the health system in terms of safety, accessibility, and efficiency by ensuring limited interaction of human in contact to devices [2], [8], [9], [22], [23], [28], [49].

2.2 Contactless and Remote Heart Rate Assessment

2.2.1 Eulerian Video Magnification (EVM) and Remote Photoplethysmography (r-PPG) for Contactless and Remote HR Assessment

Over the past decade, research has majorly focused on the implementation of remote and contactless technologies to improve and advance digital health monitoring technology [23], [26]–[28], [50]. The advancement of remote and contactless methods for HR measurement has been ongoing for the past several years, but it all started back in 2012 when two research works were published in parallel on similar concepts for a new HR remote assessment approach [25], [30]. Eulerian Video Magnification (EVM) and Remote Photoplethysmography (r-PPG) are the two main techniques, and both were founded on the similar core basis of detecting the minuscule changes of skin that occur in video due to the presence and absence of blood in the capillaries during the recovery cycle of the heart.
EVM is a novel technique that involves the spatial decomposition of video into different spatial frequency bands through the Laplacian Pyramid which is then followed by the temporal filtering of the pixel values using a custom-fit band-pass filter. After that, the filtered signal is amplified using an amplification factor to reveal the minuscule color and motion changes [25]. As a result, the filtered signal is gets amplified using an amplification factor whereas noise is reduced due to the amplification to reveal the minuscule color changes [25].

Many research works have published efforts related to direct implementation of both techniques or exploration of relevant areas in EVM or r-PPG essential to HR assessment [22], [25], [32], [52]. Both techniques were found to share similar generic implementation, which includes camera type selection, video captures, and processing steps for a successful HR estimation [53]–[55]. The processing algorithm starts with spatial processing, detecting the subject in a video, followed by selecting the ROI [25], [26], [31], [56]. Many researchers have studied methods that recognize motion by automatically tracking subjects and then selecting the exposed skin regions allowing a more adaptable ROI across the frames of video [31], [33], [57]–[59]. All the pixels in the ROI are then averaged into one measure which represents a series of spatial data of the video [1], [2], [23], [27], [28]. Early research work on VM in has found that the green pixel channels should be used for proper acquisition of a HR estimate [1], [23]. The results from spatial processing then undergo temporal processing within the time/frequency domain for detection of proper components of HR [43], [48], [58], [59]. The processing requires application of various methods such as time windowing, Fast Fourier Transform (FFT), spectrogram, or Autoregressive methods to identify the relevant frequency that matches
and correlates to the HR [62]–[68]. Furthermore, sub band filtering is applied for a focused precision based on the acceptable HR, as a normal HR can vary widely from 30 to 200 BPM. Lastly, a final processing step requires identifying the HR estimate as well as evaluating some level of confidence for the HR choice [4], [25], [63], [68], [69].

Despite the successful implementation of the algorithms of EVM/r-PPG, there has been a number of challenges that were addressed, as remote assessment solutions can be adversely affected by motion, and the early work on EVM and r-PPG warrants exploration into the effects of various factors, such as illumination, skin tone, and motion, on the performance of such assessment technologies [46], [65], [67], [70].

2.2.2 Existing Review Articles on Remote Heart Rate Assessment

Of the many works on EVM and r-PPG that have been published, a variety of works have been conducted specifically focused on remote HR assessment, resulting in a number of review articles that highlight current issues that exist, such as poor signal-to-noise ratio, lack of dynamic range, motion artifacts, ambient light interference, and tissue glare. In addition, specific factors tend to adversely affect the proper HR assessment through the remote methods, including skin thickness and skin pigmentation [46], [71].

In terms of signal-to-noise ratio (SNR), r-PPG and EVM tend to have lower signal-to-noise ratios (SNR) compared to the contact methods. A study in [72] developed a new pipeline of algorithms that combines and traces signals from camera captures specifically from the 3 color channels to detect the maximum signal to lower the effect of noise. The pipeline characterizes and identifies significance of signals through statistical and physiological measures through the camera used for video [72]. In addition to SNR, a research work explored the significance of adaptive noise cancellation methods to cancel
motion in video through studying the variations of the reflected light caused by head movements, as well as skin modeling which helped in eliminating the noise [73]. Another research work developed a multi-scale ROI model in remote HR estimate that models pulse-like features on a spatial pyramid and was shown to be highly correlated but distinct from the other features [74].

Other reviews were conducted which included comparison of non-contact methods like EVM and r-PPG with other contact methods such as ECGs, microphones, headphones, PPG, and much more. The comparison focused on the differences of both types of applications in terms of cost, robustness, immunity to noise, and technology maturation statuses. The review indicated the performance of such emote methods to perform as equal to the traditional and commonly used contact methods [75].

Another review conducted in 2018 focused on the challenging aspects of vital sign assessment of a driver within an automotive environment, covering many methods based on the optical imaging techniques including visible light and thermal imaging [76].

In 2019, a thorough review in [75] focused on the limitations common in remote vital sign technologies. The review detailed works that explored the limitations through assessing performance through distance monitoring, subject tracking techniques, and the effect of choice of camera type. Another similar review explored limitation by specifically studying vital sign assessment using RGB camera in r-PPG applications [54].

Several recent research studies specifically explored skin tone limitations that tended to persist to fail previously. The work established a new diverse r-PPG that boosted the performance of VM across a variety of darker skin tones through
incorporating physics-driven algorithms using digital image processing to eliminate the effects of dark skin regions and intensify the pulse-like signals in the video [77].

A recent paper on r-PPG established a novel denoising approach using the new Action Units, a machine vision technology method, which is alternate to the common sequential architectural processing. It was seen successful in evaluating and modeling facial expression to characterize noises based on the facial region and the facial deformation [44]. The new denoising network provided improved results in HR estimation as it both outperforms the state-of-art methods and can be easily integrated into the existing remote methods to improve HR estimation and accuracy [44].

2.3 Foundation Methods for Optical HR Assessment

EVM and r-PPG are very similar implementations and they have evolved to share similar algorithms in order to assess HR [3], [45], [64]. The remote and contactless methods require acquisition of signals from a region of skin through optical sensors such as RGB, thermal, or infrared cameras. The spatial and temporal processing require the core principles of digital image processing by transforming all the subtle color variations acquired into digital signals to be further processed to extract the signals to assess HR [22], [23], [27], [28].

The first main digital processing step is to align all the single pixel measures from each ROI of every frame of video to obtain a clear representation of a time series signal. Each sample of the time signal is an average of all pixels in the ROI and the variations across the video reflect the variation due to noise as well as correlated variation related to the HR components. Digital filters are required to eliminate unnecessary frequencies and maintain frequencies in the HR range. A digital high pass filter eliminates all
frequencies below corner frequencies and accepts otherwise, and a digital low pass filter accepts all frequencies below the maximum acceptable frequency. Transforming the time signal to frequency domain allows the identification of potential HR as peaks using Fast Fourier Transform. The largest magnitude of the FFT likely relates to the HR, as has been shown in previous publications that looked into HR.

The VM method is an essential component of this thesis as it forms the base from which the confidence framework methods extend. The base VM includes the major steps that were initially introduced in EVM for remote vital sign assessment, however, for this thesis, Video Magnifications involves the magnification of signals that relate to HR which tend to strongly be correlated across time in video. First, base VM initially starts with video captures in which all frames of a video are acquired and, similar to spatial processing, the digital image processing tools allows for the spatial selection of the region of interest (ROI) of a specific area of the facial skin for individuals in a video. After ROI selection, the average of all pixel elements in the selected ROI is computed and repeatedly for all the video frames accordingly to form a time series. Equation 1 models the main signals of the $i^{th}$ pixel in the ROI where $K$ is the constant level due to the skin color, $HR(t)$ is the variation of the color due to blood flow, and $noise(t)$ is the noise which is caused by un-wanted external factors. $K$ remains constant across video duration. $HR(t)$ will occur in all the pixels of the ROI and it will be correlated across the pixels. $Noise(t)$ is uncorrelated across the pixels with mean of 0. Averaging across $M$ pixels as in Equation 2, the correlated $HR(t)$ will be magnified as the HR signal for the whole ROI is simply $M$ copies of $HR(t)$, while the $noise(t)$ will be replaced with its mean (expected value) as uncorrelated signals will cancel out resulting in a gain $\sim 0$. The result is Equations 3 and
4 where \( HR(t) \) has been magnified compared to \( noise(t) \). The \( HR(t) \) will remain the strongest and dominant signal while the constant \( K \) will provide a DC level rather than a variation as in the case of \( HR(t) \).

\[
Pixel_i(t) = [HR(t) + noise(t) + K] \\
\text{Mean Pixel}_{ROI} = [\sum_M(HR(t) + noise(t) + K)]/M \\
\text{Mean Pixel}_{ROI} = (HR(t) + EX(noise(t) + K) \\
\text{Mean Pixel}_{ROI} = (HR(t) + K)
\] (1) (2) (3) (4)

Subsequently, as in EVM, the temporal analysis using digital filtering through a bandpass filter is used to preserve only the frequencies within the acceptable HR range and eliminate all those outside of this range. The base VM also divides the full-time signal into multiple time intervals which will be required as part of the confidence framework.

2.4 Machine Learning

Machine Learning is an essential component that can extend beyond the remote HR assessment tools and methods such as VM. The use of Machine Learning allows the resolution of issues that persist to limit the performance of VM and other remote techniques like r-PPG. Machine Learning establishes solutions as, by using Machine Learning models, it can allow the extraction of trends that guide proper identification of signals correlating to HR. Machine Learning models have been widely used recently in remote and contactless EVM/r-PPG or VM. One study developed neural network models and fed it into VM results data as well as integrating infrared light to improve the detection of the oxygen saturation levels which is also known as the \( \text{SPO}_2 \). The results of
the study were in combating the challenges due to the low illumination that was caused by the low-cost camera which resulted in an improved measure of $\text{SPO}_2$ [78], [79].

This thesis includes a section on Machine Learning to assess the performance of the proposed methods that are part of the overall confidence framework. Machine Learning will assess whether the HR predicted by VM is correct or not through classification of the measured HR from the confidence framework.

2.5 Challenges with Video Magnification

There are many limitations that can affect the performance of VM which can prevent accurate HR measurement. Motion is one factor that can directly affect HR measurement through VM because motion can cause external interference from neighbouring non-skin into the chosen skin region. In this case, the regions selected no longer reflect the true color changes as they do not solely include a constant skin region but rather variable colors from different elements such as hair, clothes, and parts of the background [1], [26], [28]. In addition to interference, motion can directly change the location of skin with respect to the light source. As a result, the reflectivity of the light causes a change the pixel colors/intensity reflected as a change in the skin region [58], [80]. Another issue with VM is that variable illumination can prevent detection of the true color changes of the skin region. For example, artificial light sources may include flicker noise that affect the skin regions which prevents identification of the true color variation of the skin region [22], [27], [28]. Skin color is also a third factor which limits VM performance, as darker skin tones can be more problematic than fair skin tones because
darker skin can limit the visibility of the color changes caused by the capillary movements from blood flow in the systole/diastole process [31], [35].

2.6 Confidence in HR Measurement

HR is a physiological parameter whose assessment tools have been a major concern in the research of advanced health solutions and technologies due to the level of uncertainty that can be introduced alongside the limitations that had been addressed of VM [1], [28], [81]. Previous work that explored limitations by proposing new and effective techniques included assessment of the measured HR by validation through ground truth HR measurement [1], [2], [22], [23], [27], [28]. However, performing HR measurement solely through a standalone remote technique can be challenging, especially with limitations which reduces the level of reliability in choosing the correct HR measure [23], [44], [81]. Confidence assessment in HR measurement through remote and contactless tools is highly essential to be able to create a decision criterion around the VM-predicted HR measure. The confidence assessment improves the reliability of the methods [82]. Many researchers specifically explored the limitations by addressing successful methods that mitigated the effects to certain extents, but VM still failed for a number of cases [27], [28], [44]. However, it is mostly that failure or success was based off of analysis that working from the ground truth HR measures. There is no evidence of previous work that explores specifically confidence assessment in the VM predicted HR. The closest parallel available is various research that has explored advanced machine vision technologies in the realm of video analysis for person tracking in order to build levels of confidence in the decision prospects and results [83].
The confidence assessment can be the initial point to incorporating the advanced computerized algorithms leading to optimized measurement of HR and improving the reliability of the VM method [82].

2.7 Spatial Domain within Video Magnification

A number of research works have explored VM in the spatial domain, which involves the ROIs being selected for HR measurement [23], [27], [74], [80]. One research work explored the effect on VM based on selected ROIs. Researchers explored the effect of the size of the ROI by dividing the full ROI into smaller ones. The method was found robust for ROI above a minimum certain size [23]. However, it constantly fails for cases of motion and when there is no masking of facial regions. The researchers discuss the required necessity to explore motion tracking solutions to improve the measurement of HR though VM [23]. Some studies proposed a Face-reader framework with algorithms to track the ROI in a video of a moving subject [56]. This method had been found successful in motion tracking to eliminate its adverse effect; however, it did not show robustness in large degrees of motion and performance degrades as the distance between the subjects and camera increases [56]. Another study explored dividing a larger ROI into many smaller ROIs where ROIs with no HR measures are considered as outliers and are eliminated from analysis. Additional work explored the impact of motion on VM applications through increasing levels of Gaussian Pyramids instead of the usual Laplacian Pyramid during the spatial decomposition [84]. This method improved VM performance but did fail to eliminate noise and is not robust in cases of sudden motions such as head nodding. The study discussed the need for ROI tracking; however, it can be computationally expensive if ROIs are large and for large video files, and it may require studying motion induced light
variation [58]. In addition to motion and its effects on VM, a skin reflection model studied and analyzed the effect of light intensity and hue on the performance of VM. The study is extended by introducing the concept of a DC signal such as combining a fixed and a non-changing background with other ROIs to analyze variations for a more reliable HR measurement. These studies assisted in understanding flicker noise and motion impacts on VM performance [85]. Previous work also explored different illumination conditions to improve the accuracy of measured HR. The work proposed combining LED and Halogen lights for an increased illumination which basically altered the original pixel values, resulting in improved measures for HR as well as increased accuracy, though it did still fail for cases of body motion [28].

2.8 Temporal Domain within Video Magnification

The temporal domain in VM focuses on the responses of VM across the video frames temporally. Several works explored limitations by proposing techniques that analyzed VM across multiple time intervals. A previous study proposed skin extraction along video stabilization techniques, a process which selects and extracts the skin from the background for every time frame followed by stabilization through feature detection, extraction, and matching along with skin detection. The techniques were applied on subjects and did show increased performance of HR detection in comparison cases when such preprocessing method was not applied [80]. Several studies have distinctly explored VM using different time windows such as 10s [86], 20s [87], and 30s [67]. Additionally, the study in [28] found that increasing the time window length up to 30 sec. improved HR measurement for subjects that showed frequent head movement; however,
interestingly, the accuracy of the measured HR reduced in the absence of head motion [28].

Chapter 3: Experimental Set-up

This chapter includes a description of and essential detail relating to all the experimental setup from tools to procedures leading to the acquisition of the full dataset as well as the initial implementation of the VM which is the core for all the methods of this thesis.

3.1 Dataset

This section provides details on the experimental set-up for data collection which led to the dataset that was used in this thesis. The dataset being used is a subset of a pre-existing dataset that consisted of a collection of RGB video recordings using a Panasonic-RGB camera as well as PPG data through a pulse oximeter (8100S, Nonin, Plymouth, MN, USA) which were obtained for all 22 subjects [1], [2], [23], [28]. The dataset was used as part of previous research studies within the same project team which explored areas and presented early work for VM as well. The data collection was performed by NRC and approved by the NRC and Carleton Research Ethics Board [1], [2], [28]. The 22 subjects that participated in the data collection were employees at NRC who provided written consent [1], [2], [23], [28]. Data collection took place during the COVID-19 pandemic year of 2020 and during the time period when social distancing enforcement measures and restrictions were mandatory [23], [27], [28]. Therefore, the limited access to the office played a major role limiting the subject group demographics, resulting in the group showing a wide range of ethnicities but a significant sex imbalance as the males
outnumbered the females. Of the 22 subjects who participated in the data collection, only 19 subjects were used for this thesis, two of whom were female and the other 17 male [1], [2], [23], [28]. The three subjects were excluded as one had hair covering the forehead during video captures, making it not convenient for testing and experimenting for VM applications. Another subject had a highly variable/abnormal HR even within the PPG sensor, and a third subject had missing data files that were required in the research work of this thesis [1], [23], [27], [28]. Appendix A.1 shows the list of the 19 subjects prepared by a former research student in which each are mapped with new numbers.

The dataset consists of video captures/recordings as well as the raw photoplethysmography (PPG) signals for each subject, and each video capture was performed under varied test conditions/cases. Conditions involved using a specific type of illumination source such as LED light, Halogen light, incandescent, fluorescent, infrared, as well as having subjects in specific state conditions (seated/standing and with/without face mask) [1], [2], [23], [27], [28]. For this thesis, however, only a subset of the full dataset is used, representing a smaller dataset pertaining to the test condition where subjects are seated and have no face mask, using a Panasonic-RGB camera, and having LED as the artificial light source (GVM-560AS2L LED) for illumination. For this test condition, Fig. 3.1 illustrates an overview of the experimental set-up for the data collection which also depicts the essential data acquisition components [28]. The PPG data was acquired through a pulse oximeter placed on the subject, and an acquisition set-up able to capture the PPG signals with a sampling frequency of 75 Hz on a LabView software [1], [23], [28].
During data collection, the subjects were asked to remain seated and still while looking straight ahead into the camera during the ~90 sec. video captures. Video captures were acquired using a Panasonic camera with a varifocal lens of 12mm, an acquisition rate of 30 Hz or 30 frames per second, and a pixel resolution of 800x600 [1], [23], [28]. The camera was placed at a distance of 2.1m from the subject and a height of 1.8m from the ground. Illumination was provided using two LED light sources enclosed in soft boxes and placed at a height of 1.9m from ground and specifically positioned at a face-on-camera angle [28]. The subjects were placed in a closed anechoic chamber allow for lighting control, which then resulted in a limited free space to place the cameras in the corresponding position. The aim was to ensure the camera is capturing all of the upper body with some background, and the corresponding height and distance worked well for all subjects.

Once all the data was collected, dataset was structured by storing and organizing the video files into categories of the several test cases. In addition to the video files, all the PPG collected data was also structured into a matrix through MATLAB (SenSmartPPG.mat file) [28]. The .mat file includes 21 entries pertaining to each subject (only Subject 22 is not included in the full PPG dataset), and, for every subject entry, a “trial” field with 10 different entries pertaining to a test condition and including 3 fields; the sampling frequency $f_s$ of value 75 time, and vals which is a 1-D vector array with all the acquired PPG data. Both the video and PPG data were stored on a 10-terabyte external-desktop drive (Seagate-Expansion).
3.2 VM implementation method

This thesis builds on a base video magnification method and part of the primary task was to design and implement the full VM algorithm to ensure results are correct for HR assessment. VM primarily focuses on video data such as the NRC dataset, so video processing is an essential component of the preliminary stages of the research. Initially video acquisition, video processing tools, and methods to properly implement VM were implemented using videos captured of myself. The work started with videos of ~80 seconds captured while lying prone to minimize body motion. The test condition was otherwise similar to the chosen test condition from the NRC dataset in terms of type of camera, state conditions, and illumination. The video captures were obtained through an RGB camera of an iPhone 8-series cellphone and placed at a distance of 30 cm ensuring the full face appears clearly on camera. No artificial light source was initially used during
those video captures except for sunlight (color temperature of 5000 Kelvin). Since this thesis relies on PPG data for ground truth measure of HR, a portable pulse oximeter (Model: BM 1000) was placed on the index finger alongside video capture. Afterwards, the video file was converted into a .mp4 file through a video player software on a laptop to be imported into MATLAB. Fig. 3.2 below summarizes the video acquisition process as well as summarizing the essential video acquisition process that will be used for the VM algorithm.

**Fig. 3.2. Experimentation and test conditions set up for video and PPG acquisition to test video processing for video magnification algorithm.**

MATLAB software includes a video processing toolbox which directly imports the video file into MATLAB for further processing. The videoreader (“filename.mp4”) takes in the .mp4 video file and creates a video object and explicitly lists all the properties related, such as the total number of frames in video, frame rate, and pixel-by-pixel resolution.
Fig. 3.3 depicts the major steps to implement VM leading towards HR assessment. The assessment initially starts by video processing to properly import all the video frames, the selection of a fixed and single ROI, computation of an average pixel measure for the ROI image which forms a stack of a time series that is represented as a time signal, that is further filtered using a bandpass filter. The filtered time signal is transformed into the frequency domain by the computation of the FFT which then undergoes post-processing through application of several mathematical computations leading to HR assessment. VM studies the variation in video and therefore it is crucial to ensure all the video frames are fully imported. As part of video processing and after the video is imported into MATLAB, all the frames are structured into a matrix, which is a 2-D array, and each entry is assigned a specific location for every single frame. Fig. 3.4 illustrates the structuring of all the frames and storing them into specific entries in the matrix. This is the first algorithmic step for the initial stages of VM which was implemented through two nested loops where the outer loop cycles through all the frames incrementally starting from the first frame (Frame 1) up to the last frame (Frame L) where L holds the total frame number.
Fig. 3.3. Block diagram for the framework of the base VM method.

Fig. 3.4. Structuring all the imported video frames of video into a 2-D matrix array.

The next step is the selection of the ROI from which HR will be assessed, which requires several stages from localization of a single ROI on a single frame and applying it
to all the frames of video. Fig. 3.5 represents all the three channels for the frame images that consists of a MxN pixel matrix that make up the image. The selection of ROI requires locating 4 points on the target skin region while ensuring it does not include any occlusion elements or background elements of the image. The selection requires several stages, first deciding on a frame for which ROI coordinates will be obtained. For this particular thesis, the 200th frame is always selected for ROI selection to avoid any initial variations at the start of the session. Figs. 3.5 and 3.6 depict the main steps for ROI selection on a single frame. Appendix B.1 shows the code stack for the algorithm to allow the user to manually select the 4 different borders of an ROI through impixel() which stores all the vertical and horizontal (xx and yy coordinates) coordinates into two separate 1-D vectors (xx’ and yy’). The ROI is a rectangular region WidthxLength where width includes the distance from [min(xx):max(xx)] and length extends a distance from [min(yy):max(yy)].

Fig. 3.5. Illustration showing the 3 Red-Green-Blue channels for the image frame in.
Fig. 3.6. Illustration of ROI selection on a single inner frame using - marker points on the forehead skin region.

After selecting an ROI, the rest of the image gets cropped out and the ROI information replaces the original frame image in the 2-D matrix. Fig. 3.7 illustrates and explains the calculation of the average pixel measure for each of the ROI frames. The boxes in green in Figs. 3.5 and 3.7 indicate that only the green channel [88] is used in VM as it was shown to provide the best VM results for HR measurement [1], [89].

Fig. 3.7. Generic flow diagram for computing average pixel for a selected single ROI.
After computing the average pixel values, the single measures for each ROI provide a time series having 30 frames per second represented as a time domain signal which then undergoes filtering to ensure processing focuses on the HR range and remove unnecessary frequencies. Fig. 3.8 depicts the workflow for the filtering process of the time domain signal. The bandpass filtering is performed using a digital Park’s McLellan filter that consists of a low pass filter and a high pass filter. First, a low pass filter is used with lower and upper corner frequencies of 3.0 Hz (180 bpm) and 3.5 Hz (210 bpm). Fig. 3.9 shows the response of the designed low pass filter with a passband error of 0.01 and a stopband ripple of 0.001. Second, a high pass filter, shown in Fig. 3.10, is designed with lower and higher corner frequencies of 0.45 Hz (27 bpm) and 0.65 Hz and (39 bpm). Fig. 3.11 shows the results after performing bandpass filtering removing unwanted frequencies from the original time signal.

**Fig. 3.8. Illustration of time-domain processing of the ROI stack involving bandpass filtering to preserve the HR signals.**
Fig. 3.9. Low pass filter using Park’s MacLellan FIR Filter with corner [3, 3.5] Hz.

Fig. 3.10. High pass filter using Park’s MacLellan FIR Filter with corner [0.45, 0.65] Hz.
Fig. 3.11. Mean Pixel value across time for a ~90 sec. video before filtering (top) and after bandpass filtering with a low pass filter (middle) followed by a high pass filter (bottom).

After bandpass filtering, the frequency domain processing part of VM application starts by transforming the time signals into frequency components as depicted in Figs. 3.12 and 3.13, showing the FFT plot results to graphically locate the largest peak, which is ~52 BPM, indicating the HR. The largest peaks are predicted as the most likely HR estimate for the individual. The testing of the algorithm requires validation with the actual HR measure which is in this case HR indicated through PPG signals from pulse oximeters.
It is noteworthy to specify that the early stages of VM implementation did not include windowing for the time signal and any padding in the full-time signal was used. However, windowing will be required in the coming steps.

The next step is testing the full algorithm for VM on the NRC dataset and estimating HR from the VM results obtained for each of the 19 NRC subjects. PPG signals are optical
sensory data and the time series of measures can be processed the same way (Fig. 3.14) as the time series of average pixels. Finally, the same VM algorithm is applied to all the PPG data to extract the actual HR. The sampling rate for the acquisition data of both PPG (75Hz) and video (30 frames per second) conforms with the Nyquist rate theorem to properly allow measurement of typical HR frequency. The normal ranges of HR all fall well below 1/2 of the sampling rates for both PPG and video acquisition data.

![Generic block diagram for VM on NRC dataset (Video and PPG data) obtained from the 19 NRC subjects.](image)

**Fig. 3.14.** Generic block diagram for VM on NRC dataset (Video and PPG data) obtained from the 19 NRC subjects.

### 3.2.1 Time Windowing Method

If VM is applied to the full video duration, then all variations across the video are combined. The measured HR from VM will combine any HR variation that occurs. The use of time windows allows for the measure of HR within specific time periods, allowing for knowledge of what the HR is and when the measure occurred. The proposed confidence framework is a set of methods that will assess the reliability of the measured
HR. It is beneficial to break the video into several time intervals by segmenting across time with a 50% overlap across the segments. Since HR is known to minimally vary within a short interval of time [2], [9], [49], [90], [91], the time segments are chosen to be 20 sec. of length, and a 50% overlap provides a fair amount of video data to exist among different segments. The choice for the 20 sec. of length provides more commonality between the adjacent windows. Fig. 3.15 depicts the main stages of VM on multiple time windows using time windowing technique. Dividing the full video into 20-sec time windows with a 50% overlap leads to a total of seven time windows. The 20-sec. segments are a finite length of a total of 600 samples points (600 video frames) for a sampling rate of 30Hz. Therefore, there tend to be discontinuities at the edges of each time segment which requires convolution with a window to taper the edges as much as possible to ensure reduced ripples and high frequencies to appear once FFT is applied [92]. For this reason, a Hamming window is applied to each time window/segment which can improve the accuracy, reduce the spectral leakage through reducing unwanted high frequency occurring around the true frequency [92]. Afterwards, the frequency domain processing involves the computation of FFT to each time window. Zero-padding is additionally applied which results in 8196 FFT points. The choice for the 8196 was essential to provide a greater spread across the frequency spectrum leading to improved estimates of the FFT results. The larger number of zero padding leads to a greater number of bins where each bin is a spacing of ~0.22 BPM ([30Hz x 60sec/min / 8196] BPM). As a result, the HR assessment can lead to more accurate HR estimates. Nevertheless, this provides a spacing of ~1/10th of the 2 BPM HR accuracy range to ensure the FFT does not limit HR assessment and validation.
3.3 Summary

This chapter introduced the application of the base video magnification (base VM) method which is the core foundation for all the presented research methods for this thesis. The implementation of the base VM included experimentation and testing of the algorithm that leads to the proper HR assessment. The testing and experimentation of the base VM initially involved self-recorded video captures of myself and was then performed on the NRC dataset which is a subset of a larger population set of video recordings and PPG data (pertaining to a specific test condition). The chapter explains the main stages of VM through the transitioning from the time domain to the frequency domain in which the time domain focuses on the video processing, ROI selection, and
time domain signal representation and filtering, and is then transformed into a frequency
domain through the use of the FFT which leads to the HR assessment to select the HR
estimate. The base VM does not always properly lead to the HR estimate due to the
limitations explained earlier in the previous chapter, and the inaccuracy of the VM results
cannot be clearly understood or explained. The following chapters of the thesis will
explore the spatial and temporal redundancies that will address the discrepancies of VM
as well as introducing the new methods, the spatial and temporal domain methods, which
will used to model the confidence framework that will lead to proper HR estimation.
Chapter 4: Confidence Framework through Spatial Domain Methods

Some sections in this chapter include content that was originally published and found in © 2023 [1].

4.1 Introduction

This section includes the application of VM in the spatial domain across multiple ROIs. This section involves the spatial redundancies through the spectral representation of the VM results, as well as evaluation of the variations seen from the distinct ROIs. Furthermore, the new spatial methods which will be used to create the overall confidence framework to help identify the correct HR are introduced, presented, and explained. Both methods involve the spatial combination through averaging across the ROIs in both the time and frequency domain. The confidence framework is built around the findings of both spatial combination methods leading to a confidence measurement that will be used to properly estimate the correct HR.

4.2 Methods

The methods section starts by introducing the spatial redundancy method of VM by introducing the application of the base VM method in the spatial domain across multiple ROIs that are selected for disparate face regions. The following sections focuses on the two new proposed methods that involve time and frequency domain algorithms for the spatial combination of the ROIs in VM.

4.2.1 Base VM Spatially Across Multiple ROIs

Fig. 4.1 depicts the algorithmic framework for the spatial redundancy method which appears to be very similar to the base VM seen for a single ROI. Due to the spatial
changes involved, iterative time domain processing is involved in which all the time domain steps of the base VM are applied to each individual ROI selected. Fig. 4.1 illustrates the spatial redundancies for multiple ROIs, and the preprocessing involves the same computations as the base VM through mean pixel averaging, time signal representation, and bandpass filtering, which are individually applied for each ROI. In fig. 4.1, ROI R1 is chosen on the forehead, R2 is chosen on the Red Cheek, R3 is chosen on the Left Cheek, and R4 is chosen on the Nose. However, ROIs can be randomly chosen as there is no rule of thumb for the order of the selected regions. The frequency domain starts by computing FFT for each ROI; however, the post-processing step (in dark orange) is a new step which requires global normalization across all the FFTs for the selected ROIs. All the FFTs are normalized based on the largest FFT value in which the largest peak across all four FFT has a maximum magnitude of 1 and all the rest FFTs are scaled accordingly. The final step requires HR assessment which allows identification and selection of the HR estimate according to the FFT results. VM identifies the largest peaks of the FFT plots for a selected ROI as the potential and correct HR for an individual.
4.2.2 Spatial Frequency Domain Averaging (SFDA) Method

This section includes the first proposed method that combines all selected facial ROIs through averaging all FFT results from base VM for improving HR estimation. The ROIs selected are different disparate face regions, and for this thesis, the regions are selected on the forehead, nose, left, and right cheeks since these are the facial regions that mostly do not get obscured and covered by external factors such as hair or clothes.

Fig. 4.2 shows the block diagram for the first proposed method to combine the four spatial regions to measure the heart rate. In this method, the VM results for the multiple ROIs are combined in the frequency domain by averaging all four FFT magnitudes resulting from the base VM method for a specific time window, resulting in a single average FFT ($FFT_{avg}$). The results of this FFT are interpreted similarly to the base Video Magnification method where each of the peaks are potential heart rates. Typically, the highest peak is chosen as the heart rate. In this work, a confidence measure is proposed that compares the relative height of the highest and second highest peak using equation 3.
that builds on the work in [27]. The minimum value of this ratio is 1 and, the larger the value, the greater the spread between the largest peak \((LP)\) and second largest peak \((SLP)\), providing a measure of how distinct the highest peak is within the spectrum.

\[
\text{FFT Ratio} = \frac{\text{Normalized Magnitude}_{LP}}{\text{Normalized Magnitude}_{SLP}} \quad (5)
\]

4.2.3 Spatial Time Domain Averaging (STDA) Method

Fig. 4.3 shows the block diagram for a second method to combine the ROI regions to assess HR that again builds on the base VM method in Fig. 4.1. In this method, the time
series average green pixel for each of the four ROIs are combined into a single time series by averaging them. The resulting time series is processed using the same windowing, pre-filtering, and FFT method. Again, the resulting FFT is plotted and normalized to the largest peak. Eqn. 1 is again used to provide a measure of confidence in the measurement through assessment of the spread between the largest and second largest peaks.

![Block diagram for heart rate measurement using spatial time domain averaging (STDA) method for a 20 sec. video. © 2023 [1]](image)

### 4.3 Results

The results section includes the spectral analysis for all the FFT plots resulting from all methods to evaluate the performance of those methods through assessment of the VM predicted HR. The results section is divided into 3 different groups, A, B, and C, and includes example subjects for the 3 different cases where VM showed good performance, weak performance, and poor performance respectively. The good performance of VM in Group A is cases where VM correctly predicted the HR across all the selected ROIs. The weak performance in Group B includes cases showing conflicting results either through
incorrectly predicting the HR for at least one ROI or due the significant presence of inaccurate results along the correctly predicted HR. The poor performance in Group C is cases where none of the selected ROIs predicted the correct HR.

4.3.1 Results for Base VM Across Multiple ROIs

4.3.1.1 Group A: Cases with Good Performance

Figs. 4.4, 4.5, and 4.6 show VM results through FFT plots for multiple ROIs where R1 is on the forehead, R2 is on the left cheek, and R3 is on the right cheek for Subject 3 of group A, for a 20 sec. of video (time window 1) and whose PPG HR is ~52 BPM. Subject 3 is an example subject where base VM predicted the correct HR is all the selected ROIs as well as showing very minimal inaccurate results, resulting in an increased confidence in the correctly predicted HR.

The FFT plots for ROIs R1, R2, and R3 showed significant and strong peaks at ~52 BPM and other peaks that are very minimal. From the three VM results, there is a high level of confidence that Subject 3’s HR is ~52 BPM. The measured HR matches the PPG HR.
Fig. 4.4. FFT plot for VM on ROI R1 (forehead) for a 20 sec. of video (window 1) for Subject 3 with PPG HR of ~52 BPM.

Fig. 4.5. FFT plot for VM on ROI R2 (left cheek) for a 20 sec. of video (window 1) for Subject 3 with PPG HR of ~52 BPM.
4.3.1.2 Group B: Cases with Weak Performance

Figs. 4.7, 4.8, and 4.9 show VM results for ROIs where R1 is chosen to be the forehead, R2 is left cheek, and R3 is nose for a 20 sec. of video (time window 4) for Subject 7 of group B and whose PPG HR is ~60 BPM. This example subject is representative of when base VM correctly predicted the HR in all the selected ROIs; however, there existed inaccurate results with the same level of significance to the correct HR which reduced the confidence around the correctly predicted HR.

ROI R1 shows a largest peak at ~61 BPM as well as other secondary peaks that are less significant. Likewise, ROI R2 shows a large peak at ~61 BPM as well as strong secondary peak at ~37 BPM and tends to have a normalized magnitude of 0.5 close to the largest peak. There also exist other smaller peaks dominating the spectrum that are clearly visible. ROI R3 (Fig. 4.9) shows a largest peak at ~61 BPM with a secondary peak strongly visible but less significant at ~37 BPM. Additionally, there exist other smaller peaks across the spectrum that are clearly visible. The three ROIs all showed the largest
peak around the same HR, which also matches the PPG HR. However, it is a bit conflicting due to the presence of secondary peaks which appeared to have large, normalized magnitudes which are somehow close to the largest peak. For this reason, there is a lower level of confidence in identifying the HR for this subject.

Fig. 4.7. FFT plot for VM on ROI R1 (forehead) for a 20 sec. of video (window 4) for Subject 7 with PPG HR of ~60 BPM.

Fig. 4.8. FFT plot for VM on ROI R2 (left cheek) for a 20 sec. of video (window 4) for Subject 7 with PPG HR of ~60 BPM.
Fig. 4.9. FFT plot for VM on ROI R3 (nose) for a 20 sec. of video (window 4) for Subject 7 with PPG HR of ~60 BPM.

Figs. 4.10, 4.11, and 4.12 are VM results for Subject 11 of Group B for 20 sec. of video (time window 5) for ROIs where R1 is chosen on the forehead, R2 on the left cheek, and R3 on the nose and whose PPG HR is ~86 BPM. This is an example subject where base VM incorrectly predicted the HR for one of the three selected ROIs in addition to having other inaccurate results dominating the actual HR which reduced confidence in the actual HR.

ROI R1 has its largest peak at ~86 BPM; however, there are multiple strong secondary peaks dominating the spectrum as well. For ROI R2, there is a large peak at ~47 BPM and as well as secondary peaks clearly visible but not as strong. ROI R3 shows its largest peak at ~86 BPM and two strong secondary peaks as well. Two ROIs indicate a HR measure of ~86 BPM except for ROI R2, resulting in low confidence levels in assigning ~86 BPM. The decision requires further exploration and warrants a confidence framework to increase reliability in the ~86 BPM which is the actual HR for Subject 11.
Fig. 4.10. FFT plot for VM on ROI R1 (forehead) for a 20 sec. of video (window 5) for Subject 11 with PPG HR of ~86 BPM.

Fig. 4.11. FFT plot for VM on ROI R2 (left cheek) for a 20 sec. of video (window 5) for Subject 11 with PPG HR of ~86 BPM.
Fig. 4.12. FFT plot for VM on ROI R3 (nose) for a 20 sec. of video (window 5) for Subject 11 with PPG HR of ~86 BPM.

4.3.1.3 Group C: Case with Poor Performance

Figs. 4.13, 4.14, 4.15, and 4.16 show VM results for four ROIs where R1 is chosen on the forehead, R2 on the left cheek, and R3 on the right cheek, and R4 on the nose for Subject 16 of Group C for 20 sec. of video (time window 3) and whose PPG HR is ~68 BPM. Subject 16 is an example subject where base VM incorrectly predicted the HR across all the selected ROIs, which resulted in a very low confidence as there is no evidence of the correct HR.

R1, R3, and R4 show largest peaks at ~40 BPM, while R2 shows its largest peak at ~52 BPM but all share many secondary peaks that are significantly large. The inconsistency of largest peaks and the presence of many secondary peaks provides weak confidence in assigning the HR. None of the peaks match the PPG HR. Nevertheless, the presence of large secondary peaks round the PPG HR warrants further analysis and requires a framework that will successfully detect the correct HR measure instead of solely relying on largest peaks directly from a single FFT plot.
Fig. 4.13. FFT plot for VM on ROI R1 (forehead) for a 20 sec. of video (window 3) for Subject 16 with PPG HR of ~68 BPM.

Fig. 4.14. FFT plot for VM on ROI R2 (left cheek) for a 20 sec. of video (window 3) for Subject 16 with PPG HR of ~68 BPM.
Fig. 4.15. FFT plot for VM on ROI R3 (right cheek) for a 20 sec. of video (window 3) for Subject 16 with PPG HR of ~68 BPM.

Fig. 4.16. FFT plot for VM on ROI R4 (nose) for a 20 sec. of video (window 3) for Subject 16 with PPG HR of ~68 BPM.

The base VM results ensure a strong and significant peak in the FFT plot at the HR estimates for all the selected ROIs. However, when VM fails to accurately measure HR, not only do multiple peaks occur in the spectrum, but they also tend to fall at varying locations across the spectrum. Such factors can essentially identify whether a predicted
HR is an estimate or an incorrect measure of HR. Combination of VM results can preserve the significant FFT values that would mainly include the correct HR estimate while eliminating unwanted and incorrect values.

Fig. 4.17 below is an example of the effect of averaging all the FFTs of all the selected ROIs on the Group B example subject in Figs. 4.7, 4.8, and 4.9, the result of which correctly selects and identifies the HR estimate. The resulting FFT spectrum shows a significant largest peak around the HR estimate which matches the actual HR from PPG for Subject 7, even though some ROIs for Subject 7 failed to show a largest peak at the correct HR on their own. Since averaging indicates the largest peak around the actual HR, this is an effective and a reliable technique which allows proper HR measurement in VM when some ROIs still fail. Nevertheless, the normalized magnitude (NM) for the largest peak after averaging tends to be significantly high in comparison to the secondary peaks, unlike the individual ROIs. The largest peak in Fig. 4.17 has shown to have a larger spread between the largest peak and a secondary peak, unlike the individual FFT plots. The larger spread between the largest and second largest peaks can allow stepping closer to define some confidence measures around the HR measures at the largest peaks. Therefore, the following section focuses on proposing a new confidence metric and different averaging techniques that will improve the HR measurement by building confidence in the measured and estimated HR.
4.3.2 Results for the SFDA Method

This section provides results for the application of the SFDA method on the video dataset for the 19 NRC subjects. Since there are some similarities in the response and results among groups of subjects, only a representative sample is provided through the spectral representation of the method’s results along the distinct FFT results for the four ROIs from the base VM method and having ROI R1 on the forehead, R2 on the Left Cheek, R3 on the right cheek, and R4 on the Nose.

Fig. 4.18 shows the FFT results from VM for ROIs R1, R2, R3, and R4 for Subject 17 of Group A who has a very light skin tone compared to other subjects, and who tended to remain still, showing very minimal to no movements through the video. This example subject is a representative of Group A, which also includes Subjects 2, 3, and 10. In fact, all these subjects have a very light skin color. This example subject is chosen to represent
cases where SFDA correctly predicts and shows similar results as the base VM having very minimal inaccurate measures.

All four ROIs have largest peaks at \(~48\) BPM with normalized magnitudes of 1. The FFT\_Ratio is \(~3.9\) for R1, \(~4.2\) for R2, \(~4.3\) for R3, and \(~4.5\) for R4. Fig. 4.19 shows the FFT plot results from the SFDA method which indicates a largest peak is at \(~48\) BPM with a FFT\_Ratio of \(~4.3\). Both base VM and SFDA indicated a largest peak at the correct HR matching the PPG HR, and therefore SFDA correctly identified the HR for Subject 17 with an FFT\_Ratio higher than most of the ROIs. Therefore, there is an increased confidence in the identified HR estimated by SFDA.

Fig. 4.18. FFT results of VM for Subject 17 for 20 sec. of video (window 5) for ROIs R1, R2, R3, and R4. The PPG HR is 47.8 BPM.
Fig. 4.19. FFT result plot of VM for the same example subject in Fig. 4.18 using SFDA method.

Fig. 4.20 shows the base VM results for selected ROIs, R1 (forehead), R2 (left cheek), R3 (right cheek), and R4 (nose) for Subject 1 who had a medium to dark skin tone and repetitively performed eye movements characterized by directing eyesight downward then followed by abrupt blinking that occurred in one of the time windows of video. This example subject is sample representative for Group B cases where VM showed and predicted the correct HR but the results included significant conflicting results as well. This example subject represents group B which includes Subjects 1, 5-7, 9, 12-16, and 19, in which the majority of subjects had medium skin tone, but very few had light skin tone, and showed some motion for throughout the video for some of the subjects. This example subject describes the case where SFDA predicts the correct HR when base VM fails for one of the ROIs; however, SFDA performance degrades in comparison to the Group A example subject (Figs. 4.18 and 4.19).

The largest peak for all four ROIs for Subject 1 is at ~66.9 BPM; however, there are secondary large peaks strongly visible as well as inconsistently appearing at multiple
locations across the HR spectrum. The FFT_Ratio is ~4.3 for R1, 1.1 for R2, ~1.3 for R3, and ~2.2 for R4. Fig. 4.21 shows the results of the SFDA method combining the results from Fig. 4.20. Similar to all ROIs, a largest peak occurs at ~66.9 BPM which matches the actual PPG HR of 66.5 BPM, and the FFT_Ratio is ~1.1. The largest peaks tend to persist around the actual HR for both VM and SFDA methods even though SFDA has an FFT_Ratio lower than most of the ROIs from the base VM method.

![FFT results of base VM for Subject 1 for ROIs R1, R2, R3, and R4 for 20 sec. of video (window 5). The PPG HR is 66.5 BPM.](image)
Fig. 4.21. FFT result plot of VM for the same example subject in Fig. 4.20 using SFDA method.

Fig. 4.22 shows the FFT results from base VM for ROIs R1, R2, R3, and R4 for Subject 19 who had a very light skin color compared to the rest of all the 19 subjects, as well as visible redness of skin indicating thin skin, and who also tended to remain very still, showing very minimal to no movement in the video. This subject is a sample representative for the cases where SFDA provided a correct prediction of the measured HR even though base VM predicted incorrect HR for some ROIs as well as having largely conflicting results, unlike cases in Group A.

This example subject is a shows a lot of conflicting results appearing in the FFT plots for the base VM which reduces the confidence in the actual HR and can even provide inconsistent HR measures across several ROIs, in which ROIs R1, R2, and R3 all have largest peaks at ~85.0 BPM but also have several secondary peaks that are significantly visible in the HR spectrum. On the other hand, R3 has a largest peak at ~62.0 BPM as well as other secondary large peaks dominating the spectrum as well. The majority of ROIs indicated the same largest peak around the same value which increases the
confidence in the HR measure of ~85.0 BPM, and, in fact, that measured HR matches the actual PPG HR of this subject. However, the results from ROI indicate a measured HR which falls far below R1, R2, and R3 and which does not agree with the actual HR of the subject. Fig. 4.23 shows the results of the SFDA method where the FFT plot shows a largest peak at ~85.0 BPM and an FFT_Ratio of 1.3. The results indicate that SFDA provided a correct HR measure for the subject as well as an FFT_Ratio similar to and not lower to any of the ROIs that showed correct HR while being higher than R3 with its incorrect HR measure.

![Graphs showing FFT results for ROIs R1, R2, R3, and R4](image)

Fig. 4.22. FFT results of VM for Subject 19 for 20 sec. of video (window 2) for ROIs R1, R2, R3, and R4. The PPG HR is 84.1 BPM.
Fig. 4.23. FFT result plot of VM for the same example subject in Fig. 4.22 using SFDA method.

Fig. 4.24 shows the FFT results from base VM for ROIs R1, R2, R3, and R4 for Subject 11 for Group C who had a light skin color and repetitively performed micromovements characterized by left-to-right head rotations throughout the video. This example subject falls under Group C cases where SFDA failed to predict the correct HR, similar to the poor performance of the base VM on some of the four selected ROIs.

In Fig. 4.24, ROI R1 has a largest peak at $\sim$82 BPM as well as a less significant secondary peak. On the contrary, ROI R2 has its largest peak at $\sim$39 BPM as well as multiple secondary peaks significantly occurring at several locations. ROI R3 shows a largest peak at $\sim$52 BPM as well as other secondary peaks dominating the spectrum. Similar to R2, R4 shows a largest peak at $\sim$39 BPM as well as other secondary peaks significantly visible across the spectrum. The FFT_Ratio is $\sim$1.8 for R1, $\sim$1.2 for R2, $\sim$1.3 for R3, and $\sim$1.1 for R4. In fact, only R1 has a largest peak at the actual HR for the subject, in addition to having the highest FFT_Ratio of the four ROIs. However, there is a reduced level of confidence since every ROI shows a different HR. There is inconsistency in the VM results and the confidence is reduced.
Fig. 4.25 shows results for SFDA method where the FFT shows a largest peak at \(~39\) BPM as well as multiple secondary peaks that are as significant as the largest peak as well as an FFT\_Ratio of \(~1.1\). In fact, SFDA provided a HR prediction that does not match the actual HR of 82.9 BPM. The conflicting results from VM and the SFDA results do not provide sufficient confidence in assigning the HR. Nevertheless, for both VM and SFDA, even in Group C cases, the HR tended to appear within the first few largest peaks.

Fig. 4.24. FFT results of VM for Subject 11 for 20 sec. of video (window 4) for ROIs R1, R2, R3, and R4. The PPG HR is 82.9 BPM.
Fig. 4.25. FFT result plot of VM for the same example subject in Fig. 4.24 using SFDA method.

Fig. 4.26 shows the FFT results from base VM for ROIs R1, R2, R3, and R4 for Subject 12 who had a medium to dark skin color and a true PPG HR of ~65.3 BPM. This example subject is taken from Group B where SFDA incorrectly predicted the HR estimate and base VM performed weakly, predicting the correct HR only for a number of ROIs.

Only R1 and R4 have largest peaks at the correct HR; however, there are secondary peaks visible across the spectrum such as a second largest peak for ROI R4. There are also two significantly large peaks for R2 in which none fall around the correct HR. R3 shows a largest peak at the incorrect HR as well. The FFT_Ratio is ~3.2 for R1, ~1.1 for R2, 2.1 for R3, and ~2 for R4. The results in Fig. 4.26 indicate correct HR in two of the four ROIs, which also indicated a higher FFT_Ratio than for ROIs with incorrect HR.
Fig. 4.26. FFT results of VM for Subject 12 for 20 sec. of video (window 4) for ROIs R1, R2, R3, and R4. The PPG HR is 65.3 BPM.

The results from the SFDA method in Fig. 4.27 indicate a largest peak at ~52 BPM which is the incorrect HR. However, there are other large secondary peaks such that a second largest peak at the correct HR at ~65 BPM. The FFT_Ratio is ~1.1 which is similar to the base VM cases that incorrectly identified HR (Fig. 4.26).
Fig. 4.2 shows the FFT results from base VM for ROIs R1, R2, R3, and R4 for Subject 18 who had a very dark skin color and repetitively looked downward. This example subject is a sample representative for Group C cases when SFDA did not predict the correct HR similar to the poor performance of base VM on the selected ROIs. Group C includes Subjects 11 and 18 only.

R1 shows a largest peak at ~47.0 BPM as well as other secondary peaks that are highly significant as well, and an FFT_Ratio of ~1.1. R2 shows a largest peak at ~37.0 BPM and another secondary peak significantly visible as well as other minimal peaks across the spectrum, and an FFT_Ratio of ~1.6. R3 shows a largest peak at ~37 BPM as well as other multiple secondary peaks that are almost as significant, and the FFT_Ratio is ~1.3. R4 shows a largest peak at ~38 BPM as well as a secondary peak that is almost equally as significant and several other less significant peaks dominating the spectrum, with an FFT_Ratio of ~1.0. The VM results incorrectly assign a HR of ~38 BPM;
however, the FFT_Ratio provides a lower level of confidence in that HR value which does not match the actual HR. The SFDA results in Fig. 4.29 show a largest peak at ~38.0 BPM similar to the incorrect VM results, with an FFT_Ratio of ~1.4. Both the SFDA and base VM results and the FFT_Ratio are very close to each other, and equally incorrect in their identification of HR.

Fig. 4.28. FFT results of VM for Subject 18 for 20 sec. of video (window 4) for ROIs R1, R2, R3, and R4. The PPG HR is 91.7 BPM.
4.3.3 Results for STDA Method

Fig. 4.30 shows the result for STDA method on the same example subject (Subject 17) as Fig. 4.18. The FFT plot shows a largest peak at ~49.0 BPM and other less significant secondary peaks and has an FFT_Ratio of ~3.0. The largest peaks for STDA match the largest peaks for VM results and the SFDA method which, in fact, match the actual HR, as well as having a higher FFT_Ratio which increases the level of confidence in the HR measure.
Fig. 4.30. FFT result plot of VM for the same example subject in Fig. 4.18 using STDA method.

Fig. 4.31 shows results for STDA method on the same example subject (Subject 1) as Fig. 4.20. The FFT plot shows a largest peak at ~66.9 BPM and other less significant secondary peaks as well as having an FFT_Ratio of ~1.3. Similar to SFDA, the STDA method did provide largest peaks at the actual HR as well as an FFT_Ratio that is higher for STDA than SFDA and other ROIs.
Fig. 4.31. FFT result plot of VM for the same example subject in Fig. 4.20 using STDA method.

Fig. 4.32 shows the result for STDA method on the same example subject as Fig. 4.22. The FFT plot shows a largest peak at ~85.2 BPM as well as other less significant secondary peaks as well as having an FFT_Ratio of ~2.2. For the STDA method, the largest peak matches the SFDA method and R1, R2, and R4 from base VM method, while having a higher FFT_Ratio, increasing the confidence in the correct HR measure which is known to match the actual PPG HR for this subject.
Fig. 4.32. FFT result plot of VM for the same example subject in Fig. 4.22 using STDA method.

Fig. 4.33 shows the result for STDA method on the same example subject (Subject 11) as Fig. 4.24. The FFT plot shows a largest peak at \( \approx 82.0 \) BPM as well as other multiple secondary peaks significantly visible across the spectrum as well as having an FFT_Ratio of \( \approx 1.1 \). In fact, only STDA and R1 of base VM only show a largest peak that matches the actual HR for Subject 11.
Fig. 4.33. FFT result plot for the same example subject in Fig. 4.24 using STDA method.

Fig. 4.34 shows the results for STDA method for the same subject example (Subject 12) as in Fig. 4.26 in which the largest peak falls at ~65 BPM which is the correct HR. The FFT_Ratio is ~2.6 which is closer to the FFT_Ratio for the cases where base VM indicated the correct HR and higher than that of SFDA which showed the incorrect HR.
Fig. 4.34. FFT result plot for the same example subject in Fig. 4.26 using STDA method.

Fig. 4.35 shows the results for the STDA method for the same example subject (Subject 18) as in Fig. 4.28, in which the largest peak is far below the actual PPG HR of ~91.7 BPM as well as multiple other secondary peaks. In fact, the third largest peak is significantly showing in the spectrum and is very close to the actual HR. However, STDA did fail to clearly indicate the correct HR similar to SFDA method, and the FFT_Ratio for STDA is ~1.6.
Fig. 4.35. FFT result plot for the same example subject in Fig. 4.28 for STDA method.

Fig. 4.36 shows example results for Subject 18 using the STDA method (B) and these can be compared to Fig. 4.27. The method (B) is another approach to STDA which involves the application of FFT while maintaining phase information unlike SFDA. These results show that the FFT using the method (B) is identical to the results in Fig. 4.26 for the SFDA method. In fact, this occurred in all subjects.
Fig. 4.36. Video magnification results for the same example in Fig 4.25 using a second approach for SFDA method (method STDA (B)).

4.3.4 Summary Results for All 19 Subjects

Table 4.1 shows the HR identified by the largest peak ($HR_{LP}$) for the SFDA and STDA methods, as well as the ratio between the normalized magnitude of the largest peak and the normalized magnitude of the second largest peak (FFT_Ratio) for all 19 subjects. From Table 4.1, 16 out of 19 subjects (not highlighted) have $HR_{LP}$ for the FDA and STDA methods within +/-0.5 BPM of each other and are also within +/- 2 BPM of the of the PPG HR for the respective subject, hence VM for both methods is indicating the correct HR.

For the cases in green, the $HR_{LP}$ for Subjects 11 and 12 varies greatly between the two methods where the STDA method indicates the correct HR for Subjects 11 and 12 and the SFDA method does not indicate the correct HR. For Subject 18, indicated in red, neither method detected the correct HR based on the largest peak.
The FFT_Ratio for all of the subjects is larger for the STDA method compared to the SFDA method, except for Subject 18 where both SFDA and STDA’s results are wrong. The mean FFT_Ratio for SFDA method is 2.3 (st. dev 1.3) compared to the STDA method with a mean of 2.8 (st. dev 1.3). The FFT_Ratio provides an indication of confidence in the result by showing the relative difference between the highest and second highest peaks and hence confidence in the choice. This ratio has increased between the SFDA and STDA methods and a paired t-test shows that this increase has a p-value of 0.0002, which is statistically significant. Of note, the FFT_Ratio for some of the more challenging subjects (motion, skin tone, etc.), such as example Subjects 11 and 18, are lower (<2) while other subjects without these challenges to VM have much higher values.

Table 4.1 also shows the detected HR and FFT_Ratio for R1 alone. Similar to the STDA method, R1 alone identified the correct HR for all subjects except for Subject 18 (red). Of note, ROI R1 alone and the STDA method indicated correct HR for Subjects 11 and 12, unlike the SFDA method. The FFT_Ratio is the same or higher for the STDA method over ROI R1 alone except for Subjects 11, 12, and 18. A paired t-test shows that this increase has a statistically significant p-value of 0.0078.
Table 4.1. Largest peak indicated HR and FFT_Ratio for SFDA and STDA methods compared to the FFT_Ratio for R1 alone and to the true subject HR via PPG, indicating improved performance in HR prediction (in green) and poor results (in red).

<table>
<thead>
<tr>
<th>Subjects</th>
<th>SFDA Method</th>
<th>STDA Method</th>
<th>ROI R1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$HR_{LP}$ (BPM)</td>
<td>FFT_Ratio</td>
<td>$HR_{LP}$ (BPM)</td>
</tr>
<tr>
<td>1</td>
<td>66.9</td>
<td>1.2</td>
<td>66.9</td>
</tr>
<tr>
<td>2</td>
<td>63.2</td>
<td>3.1</td>
<td>63.2</td>
</tr>
<tr>
<td>3</td>
<td>51.6</td>
<td>4.2</td>
<td>51.8</td>
</tr>
<tr>
<td>4</td>
<td>81.5</td>
<td>1.1</td>
<td>81.5</td>
</tr>
<tr>
<td>5</td>
<td>73.4</td>
<td>1.5</td>
<td>73.4</td>
</tr>
<tr>
<td>6</td>
<td>77.1</td>
<td>2.3</td>
<td>77.1</td>
</tr>
<tr>
<td>7</td>
<td>61.0</td>
<td>1.6</td>
<td>61.0</td>
</tr>
<tr>
<td>8</td>
<td>58.6</td>
<td>4.3</td>
<td>58.6</td>
</tr>
<tr>
<td>9</td>
<td>73.1</td>
<td>2.1</td>
<td>73.1</td>
</tr>
<tr>
<td>10</td>
<td>50.5</td>
<td>4.1</td>
<td>50.5</td>
</tr>
<tr>
<td>11</td>
<td>39.0</td>
<td>1.1</td>
<td>83.0</td>
</tr>
<tr>
<td>12</td>
<td>48.0</td>
<td>1.1</td>
<td>66.3</td>
</tr>
<tr>
<td>13</td>
<td>69.2</td>
<td>1.4</td>
<td>68.9</td>
</tr>
<tr>
<td>14</td>
<td>74.7</td>
<td>4.1</td>
<td>74.4</td>
</tr>
<tr>
<td>15</td>
<td>68.7</td>
<td>1.8</td>
<td>68.7</td>
</tr>
<tr>
<td>16</td>
<td>75.1</td>
<td>1.3</td>
<td>75.3</td>
</tr>
<tr>
<td>17</td>
<td>48.9</td>
<td>4.3</td>
<td>48.9</td>
</tr>
<tr>
<td>18</td>
<td>37.0</td>
<td>1.4</td>
<td>37.0</td>
</tr>
<tr>
<td>19</td>
<td>85.2</td>
<td>1.3</td>
<td>85.2</td>
</tr>
<tr>
<td>Mean</td>
<td>2.3</td>
<td>2.8</td>
<td>2.3</td>
</tr>
<tr>
<td>St Dev</td>
<td>1.3</td>
<td>1.3</td>
<td>1.1</td>
</tr>
</tbody>
</table>

4.4 Discussion

The results provided in Figs. 4.18 through 4.35 and Table 4.1 show that both averaging methods STDA and SFDA are effective techniques to combine the information from multiple spatial regions, leading to a HR estimation using VM. The STDA method shows better improvement in the assessment of HR as it was able to provide a clear
indication of the correct HR for 18 of the 19 subjects, whereas the SFDA method only did this for 16 of the 19, while individual ROIs had lower and conflicting performance.

The similarity of the results for the SFDA method (B) to the STDA method demonstrates the importance of maintaining phase information within the calculations, as both include phase. The STDA method is computationally simpler as it only requires a single FFT calculation. Improved results associated with phase inclusion likely indicate that noise, such as that due to motion and light, can have counterviewing (phase shifted) effects. For instance, movement can increase light on one ROI while reducing light on the other ROI. According to the STDA method results, the averaging technique reduced the effects of conflicting signals such as movement. It must be noted that the ROIs that are combined in the SFDA method (B) or the STDA method need to be close to each other on the body so that the HR itself is not phase shifted between them. This is a result of differing pulse transit time due to distance away from the heart. © 2023 [1]

The results shown in Table 4.1 demonstrate the improvement in the FFT_Ratio between the SFDA and STDA methods. This is further validated using the statistical test that showed a significant difference between both methods, indicating that the STDA method provides more spread between the highest and next highest peaks. This ratio also provides some indication of the confidence in the result as, for many subjects, the ratio was well above 2, while, for others, such as example Subjects 11 and 18, the result was less than 2. These later cases and others with lower FFT_Ratio provide an indication of reduced confidence that the highest peak is actually the correct HR. These subjects are more challenging with more movements and/or darker skin tone. © 2023 [1]
Although the STDA method is shown to be an effective technique to correctly measure HR, it did not provide the correct indication based on the largest peak for all subjects. Previous work [28] failed to find the correct HR for Subject 18, as it did not appear as any peak in the FFTs. The presence of a peak near the actual HR as the second largest peak warrants additional exploration to understand if the STDA method can be combined with other methods to further improve the results so that the correct HR is chosen. For example, the temporal domain could be explored where VM results for the adjacent time windows are compared to identify similar predicted HR. Finally, this paper provides a proposed method to use diverse spatial regions to improve HR assessment and to also provide a measure of confidence in the HR measure. © 2023 [1]
Chapter 5: A Confidence Framework through Temporal Domain Methods

The majority of the contents of this chapter are originally found in © 2023 [2].

5.1 Introduction

This chapter demonstrates the necessary methods to implement the confidence framework for HR estimation through temporal domain processing. The section explores VM results across multiple time windows leading to temporal redundancy that will identify the issues in the VM performance that occur temporally throughout the video duration. The next section will introduce and present the temporal domain method that combines VM results from overlapping and adjacent time windows, leading to confidence measures that will model the temporal confidence framework to improve HR estimation.

5.2 Methods

5.2.1 Application of Base VM Temporally Across Multiple Time Windows

Fig. 5.1 depicts the flow diagram of VM to measure HR across multiple time windows of a video. In this paper, the methods were tested on ~90 second video recordings [23]. After importing the raw videos into MATLAB [1], the middle 80 seconds of video were selected for analysis (2401 video frames). Next, a single ROI on the forehead is chosen as it provides the most reliable HR measure and the green-channel pixels for each frame were averaged together. The resulting time-series of green pixel values was then filtered for the HR range [1], [22] using an FIR bandpass filter (pass band corners: 0.65Hz (39 bpm) and 3Hz (180 bpm); stopband corners: 0.45Hz (27 bpm) and 3.5Hz (210 bpm)). Twenty (20) second time windows with 50% overlap were segmented
and processed with a Hamming window and zero padding to calculate 8192-point FFTs. Each of the seven time windows provided an FFT result and the seven results were normalized so that the largest peak across all seven was set to magnitude 1 [1]. The frequencies of the largest and second largest peaks in each FFT were located, with the frequency of the largest peak selected as the estimated HR [1]. The estimated HR was considered correct if it was within ±2 BPM of the PPG HR for the same window [93]. Dr. Neil Thomas and Dr. Frank Knoefel are two physician who were part of the research team and provided proper guidance and feedback on medical aspects. From a medical judgement perspective, Dr. Thomas and Dr. Frank indicated that a difference of 4 BPM would not have a significant impact on clinical decision. Therefore, +/-2 BPM of the true HR would still be considered an acceptable and correct HR measure by VM.

Fig. 5.2 illustrates the process for calculating a confidence metric for each time window based on the ratio of the largest peak to the second largest peak (FFT_Ratio) [1]. The FFT_Ratio provides a measure of the confidence in a given largest peak that it holds the correct HR, and this thesis considers how this measure can provide confidence that an assessed value is correct or invalid. For each subject, the average FFT_Ratio (AvgFFT_Ratio) was also computed over all 7 windows.
5.2.2 Temporal FFT Averaging (TFA) Method

Fig. 5.3 illustrates the new method, Temporal FFT Averaging (TFA), that combines the initial FFT results from consecutive overlapping time windows to generate a new set of FFTs for HR estimation. Normally, a person’s HR tends to vary over time, and it is expected that adjacent and overlapping time windows will have a very similar but not necessarily identical HR. For each time window, the calculation of a new FFT \( (FFT_{\text{new}}) \) is derived from the original FFTs from the preceding, current, and subsequent time windows using a nested loop algorithm (Equation 4).
This new technique calculates $FFT_{w, \text{new}}$ for time window $w$ ($FFT_w$) where the new prediction for a given frequency is the average of the maximum for the prior ($FFT_{w-1}$), current ($FFT_w$) and next ($FFT_{w+1}$) windows within a range ($R_m$) of ±m BPM (m = 1, 2, or 3). The result is a new FFT for each time window $w$ that combines the adjacent windows while allowing for some window-to-window HR changes. Similar to the base VM method, the largest peak in $FFT_{w, \text{new}}$ is considered the most likely candidate for the HR for that time window (Equation 5). One effect of the averaging method is that FFT peaks will naturally have flat tops based on the ±m BPM range and the HR is 1-D vector array with all the frequencies for the largest peaks, and the $HR_{LP}$ is the middle frequency of HR. Given the sampling rate of 30 Hz and an FFT length (bin) of a total of 8192, the values of k for ±1, 2, 3 BPM are 4, 9, and 14 FFT bins respectively. Therefore, all frequencies within the range of ±4, 9, and 14 bins of k are combined, and the largest value is stored in k. The second largest peak in $FFT_{w, \text{new}}$ is determined to allow for the calculation of the $FFT_{Ratio}$ and averaging over all five windows will compute $AVG_{FFT_{Ratio}}$ (Equation 6). © 2023 [2]
Fig. 5.3. Block Diagram for HR Measurement using TFA method. © 2023 [2]

\[
FFT_{w_{\text{new}}}[k] = \left( \sum_{w=2}^{6} \sum_{d=-1}^{1} \sum_{k = \text{bin}-1}^{L-\text{bin}} \max(FFT_{w+d}[k-R_{m}:k+R_{m}]) \right) / 3 \quad (4)
\]

\[
HR_{LP} = HR\left( \max(FFT_{w_{\text{new}}}, 1) \right) / 2 \quad (5)
\]

\[
AVG_{FFT_{\text{Ratio}}} = \frac{\sum_{i=1}^{5} FFT_{\text{Ratio}}}{5} \quad (6)
\]

5.3 Results

5.3.1 Results for Base VM Method Temporally Across Multiple Time Windows

This section includes the results of the application of base VM on the multiple time windows. The results involve the sample representative of the FFT plots for a three overlapping and consecutive time windows for a specific video. The results will provide temporal redundancy and highlight the issues in the performance of VM as time progresses along the video. The results section is divided into groups A, B, and C, where each group includes example subjects that represent a specific case of base VM performance on the multiple time windows.
5.3.1.1 Group A: Cases with Good Performance

Fig. 5.4 shows the base VM results for Subject 10 of Group A when VM is applied on a single ROI for time windows 3, 4, and 5, and the true PPG HR for each of the three time windows is ~48 BPM. This example subject is a sample representative of the cases where base VM performed well predicting the HR estimate as well as having high levels of confidence in the predicted correct HR.

All the three time windows clearly indicate a significantly strong largest peak at ~48 BPM which matches the actual HR. From the results, there is a strong confidence in assigning ~48 BPM as the correct and true HR measure for Subject 10 for the three time windows.

Fig. 5.4. VM results on a single ROI (forehead) for Subject 10 for time windows 3, 4, and 5 for with respective PPG HR of ~48 BPM.
5.3.1.2 **Group B: Cases with Weak Performance**

Fig. 5.5 shows the VM results for Subject 4 when VM is applied on a single ROI for time windows 1, 2, and 3, and the true PPG HR for each of the three time windows is ~81 BPM. This example subject is a sample representative for Group B where base VM performed weakly, incorrectly predicting the HR for a specific time window as well as showing many conflicting measures in the time window that does include the correctly predicted HR. This is a special case, as time window 4 included a massive motion causing by yawning of the subject which was not seen in any of the other subjects.

Time windows 1 and 2 show largest peaks at ~81 BPM with other secondary peaks that are significant as well; however, window 3 shows a largest peak at 60 BPM as well as other secondary peaks. Only time windows 1 and 2 show a HR that matches the true PPG HR. The presence of large secondary peaks, as well as inconsistency in the large peak for time window 3, presents low levels of confidence in assigning HR. In fact, the time windows overlap at 10 sec. which is expected to facilitate consistent HR results, but this is not the case here. Therefore, there should be a framework to increase the level of confidence to accept the correct HR, which is ~81 BPM for this case, and reject the inaccurate results.
Fig. 5.5. VM results on a single ROI (forehead) for Subject 4 for time windows 1, 2, and 3 with respective PPG HR of ~81 BPM.

5.3.1.3 Group C: Cases with Poor Performance

Fig. 5.6 shows a case where VM presents more conflicting results which make it very hard to predict and build confidence around the actual measure of HR for the subject. The FFT plots indicate more than one strong peak for a given time window, leading to a reduced level of confidence in the predicted HR estimates.

The results in Fig. 5.6 are for Subject 13 of Group C for a single ROI for time windows 4, 5, and 6, whose PPG HR is ~68 for all three time windows. This is an example subject for cases where base VM performed poorly, showing many inaccurate results that are as significant measures as the actual HR, leading to reduced confidence in the correct HR measure.

For time window 4, there are two large peaks with a largest peak slightly greater at ~57 BPM and another secondary large peak at ~68 BPM. Time window 5 shows a largest
peak at ~68 BPM and a secondary peak strongly visible at ~50 BPM. For time window 6, there are two main large peaks at ~57 BPM and ~60 BPM. For all three time windows, the correct HR appears within the first few largest peaks; however, the presence of other strong and significant peaks at HR around 50 BPM makes it harder to decide the correct HR.

Fig. 5.6. VM results on a single ROI (forehead) for Subject 13 for time windows 4, 5, and 6 with respective PPG HR of ~68 BPM.

5.3.2 Results for Temporal FFT Averaging (TFA) Method

This section includes results from the base VM method and the TFA method on the 19 NRC subjects. The subjects were divided into three groups based on the performance of the base VM (Method 1) providing a most likely HR estimate. Group A had near-
perfect performance, Group B had mixed performance, while Group C had very poor performance and, in some cases, never provided a correct estimate. The categorizing of subjects into groups will help in evaluating the performance of the new method, the TFA method through the analysis of the differences in the responses compared to the base VM. In fact, this section will include Figs. of the FFT plots that result from both the base VM method and the TFA method (TFA) for the choice of ±1, 2, and 3 BPM for the three overlapping time windows. The Tables provide the results for Method 1 and the three TFA cases (m = 1, 2, and 3). The tables show the HR at largest and second largest peak ($HR_{LP}$ and $HR_{SLP}$) and their respective normalized magnitudes ($NM_{LP}$ and $NM_{SLP}$), the FFT_Ratio and its average, as well as a last column showing the PPG HR for each time window. All the windows with $HR_{LP}$ within ±2 BPM of PPG HR are marked in green in the tables.

5.3.2.1 Group A: Cases with Good Performance

Group A consists of one female subject (Subject 2), and five males (Subjects 3, 8, 10, 14, and 15) where Subjects 2, 3, 8, 10, and 14 had a fair-colored skin tone and Subject 15 had a darker skin tone. Generally, subjects in Group A had very minimal head movements that could minimally affect the ROI location frame to frame. The one exception was Subject 14, who showed abrupt lower head and neck movements affected by swallowing motion which occurred at the specific time interval within windows 3, 4, and 5. Group A subjects showed a HR within ±2 BPM of the PPG HR for all the five inner time windows (2 through 6) using the base VM Method 1 as well as for TFA method. The internal five time windows are only considered as the TFA method requires a prior and post window to be calculated.
Fig. 5.7 shows example FFT results for Method 1 for windows 2, 3, and 4 for Subject 2, while Fig. 5.8 shows example results for TFA for the same subject for the same windows. Fig. 5.7 indicates a single largest peak around almost the same $HR_{LP}$ that is summarized in Table 5.1. Various secondary peaks occur but tend to be minimal in comparison to the largest one. The consistency of the largest peak around the same HR value increases confidence that $HR_{LP}$ is the correct HR. In fact, all three $HR_{LP}$ values fall within ± 2 BPM of the PPG HR for the respective windows. For TFA case 2 (± 2 BPM) in Fig. 5.8, the plots continue to show the dominant peak, while some secondary peaks are reduced or are merged within the largest peak and therefore disappear as distinct peaks. Fig. 5.8 also shows the flatter top for the largest peak caused by the TFA method, and the choice of the predicted HR is based on the center HR value. The predicted $HR_{LP}$ is within the ± 2 BPM of the PPG HR for those three windows. © 2023 [2]
Results in Table 5.1 summarize the results for Subject 2 for Method 1 and all the three cases of TFA, and the FFT_Ratio across all windows, with only one exception, is greater than 2. The FFT_Ratio using TFA is generally larger than in Method 1 with the case 3 of ± 3 BPM resulting in a largest $AVG_{FFT.Ratio}$. This implies that all secondary peak values for case 3 are reduced.
Table 5.1. Largest and second largest peak HR with normalized magnitude, and FFT_Ratio for method 1 and TFA for HR within ±1 BPM (case 1), ±2 BPM (case 2), ±3 BPM (case 3) for Subject 2, compared to PPG HR. © 2023 [2]

<table>
<thead>
<tr>
<th>Window</th>
<th>Method 1</th>
<th>Case 1: TFA (±/− 1 BPM)</th>
<th>Case 2: TFA (±/− 2 BPM)</th>
<th>Case 3: TFA (±/− 3 BPM)</th>
<th>PPG HR (BPM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR_{LP}</td>
<td>NM_{LP}</td>
<td>HR_{SLP}</td>
<td>NM_{SLP}</td>
<td>FFT_Ratio</td>
</tr>
<tr>
<td>2</td>
<td>65</td>
<td>0.9</td>
<td>54.9</td>
<td>0.5</td>
<td>1.8</td>
</tr>
<tr>
<td>3</td>
<td>62.6</td>
<td>0.9</td>
<td>127</td>
<td>0.2</td>
<td>4.5</td>
</tr>
<tr>
<td>4</td>
<td>63.9</td>
<td>0.9</td>
<td>51.8</td>
<td>0.3</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>63.5</td>
<td>0.9</td>
<td>50.9</td>
<td>0.4</td>
<td>2.3</td>
</tr>
<tr>
<td>6</td>
<td>64.1</td>
<td>1</td>
<td>53.3</td>
<td>0.4</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>AVG_{FFT_Ratio}</td>
<td>2.7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The persistence and consistency of a significant largest peak across windows 2 through 6, as well as the increase in FFT_Ratio and AVG_{FFT_Ratio}, increases the confidence that the HR_{LP} from Method 1 is the correct HR for Subject 2. In fact, all the HR_{LP} of all overlapping windows in both Method 1 and TFA fall within the acceptable range of the PPG HR. © 2023 [2]

5.3.2.2 Group B: Cases with Weak Performance

Group B includes example subjects for the cases when the base VM did not identify the correct HR for at least one window but did show an improvement in HR prediction after applying the TFA method. Group B, which was all male, includes Subjects 1, 5, 7, 9, 12, 13, and 17. Figs. 5.9 and 5.10 and Table 5.2 shows the results for Subject 1s who had a darker skin tone than the rest in this group, as well has eye movements characterized
by constantly directing eyesight downward and then followed by abrupt blinking that occurred within time windows 3, 4, and 5.

Fig. 5.9. Example Video Magnification results for Method 1 for Subject 1 (Group B) for three 20-second time windows of video using the region of interest shown in Fig. 5.1. The true PPG HR for windows 3, 4, and 5 are ~65.0, ~67.0, and ~66.0 BPM respectively. © 2023 [2]

Fig. 5.9 shows how the time windows have a largest peak and as well as other secondary peaks that are larger than observed in Group A. Windows 3 and 5 each have a largest peak matching the PPG HR, while the window 4 largest peak is at a much lower frequency and the secondary peak is close to the correct HR. Fig. 5.10 shows the application of TFA, after which each window has a largest peak around the same HR, which aligns with the PPG HR. In Table 5.2, all three cases of TFA show a more consistent $HR_{LP}$ in comparison to Method 1, with the third case providing correct predictions for all windows. In these cases, the FFT_Ratio values are generally lower than those observed for Group A.
Fig. 5.10. Example Video Magnification results for TFA case 2 for Subject 1 (Group B) for three 20-second time windows of video using the region of interest shown in Fig. 5.1. The true PPG HR for windows 3, 4, and 5 are ~65.0, ~67.0, and ~66.0 BPM respectively. © 2023 [2]

Table 5.2. Largest and second largest peak indicated HR with normalized magnitude, and FFT_Ratio for Method 1 and TFA for HR within ± 1 BPM (case 1), ± 2 BPM (case 2), ± 3 BPM (case 3) for Subject 1, and compared to PPG HR. © 2023 [2]

<table>
<thead>
<tr>
<th>Window</th>
<th>Method 1</th>
<th>Case 1: TFA ( +/- 1 BPM)</th>
<th>Case 2: TFA ( +/- 2 BPM)</th>
<th>Case 3: TFA ( +/- 3 BPM)</th>
<th>PPG HR (BPM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR_Lp</td>
<td>NM_Lp</td>
<td>HR_SLP</td>
<td>NM_SLP</td>
<td>FFT_Ratio</td>
</tr>
<tr>
<td>1</td>
<td>66.5</td>
<td>0.6</td>
<td>34.7</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>65.6</td>
<td>0.7</td>
<td>120</td>
<td>0.4</td>
<td>1.9</td>
</tr>
<tr>
<td>3</td>
<td>42.6</td>
<td>0.7</td>
<td>50.1</td>
<td>0.7</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>68.5</td>
<td>0.7</td>
<td>55.1</td>
<td>0.4</td>
<td>1.6</td>
</tr>
<tr>
<td>5</td>
<td>66.9</td>
<td>0.9</td>
<td>36.5</td>
<td>0.8</td>
<td>1.1</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AVG FFT_Ratio 1.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Window</th>
<th>Method 1</th>
<th>Case 1: TFA ( +/- 1 BPM)</th>
<th>Case 2: TFA ( +/- 2 BPM)</th>
<th>Case 3: TFA ( +/- 3 BPM)</th>
<th>PPG HR (BPM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR_Lp</td>
<td>NM_Lp</td>
<td>HR_SLP</td>
<td>NM_SLP</td>
<td>FFT_Ratio</td>
</tr>
<tr>
<td>1</td>
<td>66.5</td>
<td>0.7</td>
<td>41</td>
<td>0.5</td>
<td>1.3</td>
</tr>
<tr>
<td>2</td>
<td>67.2</td>
<td>0.6</td>
<td>33.6</td>
<td>0.4</td>
<td>1.4</td>
</tr>
<tr>
<td>3</td>
<td>67.2</td>
<td>0.7</td>
<td>55.8</td>
<td>0.5</td>
<td>1.5</td>
</tr>
<tr>
<td>4</td>
<td>67.8</td>
<td>0.7</td>
<td>34.3</td>
<td>0.6</td>
<td>1.3</td>
</tr>
<tr>
<td>5</td>
<td>67.6</td>
<td>0.7</td>
<td>36.7</td>
<td>0.5</td>
<td>1.5</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>AVG FFT_Ratio 1.4</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
</tr>
</tbody>
</table>
5.3.2.3  **Group C: Cases with Poor Performance**

Group C includes Subjects 6, 11, 16, and 18 where Subject 11 was female, and Subject 18 had a darker skin tone. Figs. 5.11 and 5.12 and Table 5.3 show the example results for Subject 16, who had head and eye movements across multiple time windows of video. In Fig. 5.11, all three overlapping windows show significantly large secondary peaks in addition to the largest peak. Windows 4 and 6 both show largest peaks around the same predicted HR value that do not match the PPG rate, and have very large secondary peaks. Window 5 shows a largest peak at the PPG predicted HR. With such different window to window results, there is no confidence as to which HR is chosen as the correct HR. In Fig. 5.12, after the application of TFA, all three windows show a central largest peak around the same HR value; however, there are various large secondary peaks as well. © 2023 [2]

Table 5.3 shows the low FFT_Ratio generally across Method 1 and the three TFA cases and also the minimal improvement in these measures through application of TFA. The range of FFT_Ratios is similar between Group B and C. It is also noted that, when TFA is applied, the FFT_Ratio is generally decreased. © 2023 [2]
Fig. 5.1. Example Video Magnification results for Method 1 for Subject 16 (Group C) for three 20-second time windows of video using the region of interest shown in Fig. 5.1. The true PPG HR is 71.9 BPM for window 4, 77.4 BPM for window 5, and 73.1 BPM for window 6.
Fig. 5.12. Example Video Magnification results for TFA case 2 for Subject 16 (Group C) for three 20-second time windows of video using the region of interest shown in Fig. 5.1. The true PPG HR is 71.9 BPM for window 4, 77.4 BPM for window 5, and 73.1 BPM for window 6. © 2023 [2]

Table 5.3. Largest and second largest peak indicated HR with normalized magnitude, and FFT_Ratio for Method 1 and TFA for HR within ±1 BPM (case1), within ±2 BPM (case 2), within ±3 BPM (case 3) for Subject 16, compared to PPG HR. © 2023 [2]

<table>
<thead>
<tr>
<th>Window</th>
<th>Method 1</th>
<th>Case 1: TFA ( +/- 1 BPM)</th>
<th>Case 2: TFA ( +/- 2 BPM)</th>
<th>Case 3: TFA ( +/- 3 BPM)</th>
<th>PPG HR (BPM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$HR_{LP}$</td>
<td>$NM_{LP}$</td>
<td>$HR_{SLP}$</td>
<td>$NM_{SLP}$</td>
<td>FFT_Ratio</td>
</tr>
<tr>
<td>2</td>
<td>39.5</td>
<td>0.4</td>
<td>71.6</td>
<td>0.2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>38.9</td>
<td>0.5</td>
<td>67.6</td>
<td>0.6</td>
<td>1.9</td>
</tr>
<tr>
<td>4</td>
<td>39.3</td>
<td>0.5</td>
<td>74</td>
<td>0.4</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>76.6</td>
<td>0.6</td>
<td>67.9</td>
<td>0.6</td>
<td>1.1</td>
</tr>
<tr>
<td>6</td>
<td>38.6</td>
<td>0.3</td>
<td>57.8</td>
<td>0.2</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>AVG FFT_Ratio</td>
<td>1.6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For Window 4:

- Network:** 20.7
- LSTM:** 20.7
- CNN:** 20.7
- PTP:** 20.7
- PPG:** 20.7

For Window 5:

- Network:** 20.7
- LSTM:** 20.7
- CNN:** 20.7
- PTP:** 20.7
- PPG:** 20.7

For Window 6:

- Network:** 20.7
- LSTM:** 20.7
- CNN:** 20.7
- PTP:** 20.7
- PPG:** 20.7
5.3.2.4 Summary of Combined Results

Table 5.4 shows the combined results from all 19 subjects, separated by group, for both Method 1 and TFA cases, reporting the number of windows (out of five) that predicted HR correctly (VM HR within ±2 BPM of PPG HR), along with the minimum, maximum, and average FFT_Ratio.

Group A shows that for all cases of TFA, the $AVG_{FFT\_Ratio}$, $min_{FFT\_Ratio}$, and $max_{FFT\_Ratio}$ is generally higher than for Method 1 with a few exceptions in single measures. The standard deviation of the FFT_Ratio for TFA case 3 is also generally reduced across the subjects.

For Group B, there is some improvement through the application of TFA, allowing one or two more windows to be predicted correctly compared to Method 1. The FFT_Ratio appears to be lower for the subjects with poorer performance and, similar to Group A, higher for subjects with higher prediction performance.

For Group C, the $min_{FFT\_Ratio}$ generally alternates between 1 and 1.1, which is significantly lower than those in Groups A and B. Also, Group C shows that when Method 1 yielded a correct HR window count of zero, none of the cases for TFA resulted in an increase in the window count.

Finally, the percentage of correct windows increased from 61% up to 68% for TFA case 3, which indicates an improvement through temporal averaging to provide a better HR prediction.
Table 5.4. Minimum, maximum, average, and standard deviation for the FFT_Ratio across windows 2 through 6 for Method 1 and the three cases for TFA for the choice of ± m BPM for all 19 subjects among Groups A, B, and C, and the percentage of windows for HR within ± 2BPM. © 2023 [2]

<table>
<thead>
<tr>
<th>Grp.</th>
<th>Subject</th>
<th>Method 1</th>
<th>TFA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Window</td>
<td>minFFT_Ratio</td>
<td>maxFFT_Ratio</td>
</tr>
<tr>
<td>A</td>
<td>61%</td>
<td>1.8 4.5 2.7 5 2 4 0.8 5 2.6 4.9 4.1 0.9 5 4.5 5.0 4.6 0.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>67%</td>
<td>1.3 3.6 2.9 5 2.8 3.7 3.1 0.4 5 2.6 3.7 3.0 0.5 5 2.6 3.7 3.1 0.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>68%</td>
<td>1.4 3.6 2.3 5 1.7 3.9 2.6 0.9 5 1.7 3.9 2.5 0.9 5 1.7 3.7 2.5 0.8</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1%</td>
<td>2 1 1.9 1.3 0.4 5 1.3 1.7 1.5 0.2 4 1.3 1.5 1.4 0.1 5 1.2 1.4 1.3 0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2%</td>
<td>1.4 1.9 1.3 5 1.2 1.5 1.3 0.1 5 1.3 1.4 1.3 0.1 5 1.2 1.2 1.3 0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4%</td>
<td>1 4.5 2.2 1.4 5 1.5 2.3 1.9 0.7 5 1.4 2.5 1.9 0.4 5 1.3 2.3 1.9 0.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3%</td>
<td>1 1.5 1.2 0.4 4 1.1 1.5 1.2 0.2 4 1.3 1.3 1.2 0.1 4 1.3 1.3 1.2 0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13%</td>
<td>2 1 1.2 0.4 3 1 1.5 1.3 0.2 3 1.4 1.2 0.1 2 1.4 1.1 0.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12%</td>
<td>4 1.7 4.3 2.8 1.0 4 2 2.2 4.2 3.4 0.9 4 2 3.7 2.9 0.7 4 2 4.5 2.7 1.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17%</td>
<td>4 1.7 4.0 3.2 0.9 4 3 4.1 3.6 0.4 4 3 4.1 3.6 0.4 4 3.5 3.9 3.6 0.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19%</td>
<td>1 2.3 1.3 0.5 2 1.2 1.7 1.4 0.2 2 1.2 1.5 1.3 0.1 2 1.1 1.4 1.3 0.2</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>6%</td>
<td>1 1.1 2.4 2.0 0.5 1 1.1 2.7 1.6 0.7 1 1 3.2 1.8 1 2 1 2.3 1.7 0.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11%</td>
<td>2 1 1.4 1.1 0.2 2 1 1.1 1.1 0.1 2 1.0 1.1 1.1 0.1 2 1 1.1 1.1 0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16%</td>
<td>1 1.0 1.9 1.6 0.4 0 1.0 1.9 1.3 0.4 0 1 1.7 1.2 0.3 0 1 1.7 1.2 0.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18%</td>
<td>0 1 2 1.4 0.4 0 1 1.3 1.1 0.1 0 1 1.2 1.4 0.4 0 1 1.1 1.1 0.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4%</td>
<td>0 1.0 1.8 1.2 0.3 0 1.0 1.6 1.3 0.3 0 1.0 1.6 1.3 0.3 0 1.0 1.6 1.3 0.3</td>
<td></td>
</tr>
</tbody>
</table>

5.4 Discussion

The results have shown that the TFA method improved the VM estimation of HR for the subjects in Group B, while replicating the base VM performance for subjects in Group A. Both methods yielded poor results for Group C. Overall, the higher values for
the FFT_Ratio with the TFA method improved confidence in the choice of a given HR estimation. For Group B cases when there was mixed performance, the FFT_Ratio confidence metric is lower than in Group A. For the very poor performance cases as in Group C, the FFT_Ratio shows a minimal change across the TFA cases when compared to Method 1 and much lower values than observed for Group A. Groups B and C include cases with darker skin tones and repetitive movement. The results show that TFA improved HR measurement by increasing the confidence in the HR estimate. This is a novel technique that improves VM performance by rejecting incorrect HR values due to challenges from illumination, skin color, and motion. © 2023 [2]

Future work should combine temporal and spatial methods for the combination of VM results using multiple ROIs and multiple time windows. The work should explore methods to fuse these results and to also use the confidence measures, changes in confidence measures between methods, and changes in predicted HR between the methods to build an overall more precise HR prediction and confidence measure. © 2023 [2]
Chapter 6: Confidence Assessment of VM HR through Machine Learning

This chapter focuses on the implementation of Machine Learning models/algorithms combining the results from the spatial domain (STDA) and the temporal domain TFA that will allow assessment of whether the VM-predicted HR is correctly classified as a HR estimate, or an incorrect measure which is an invalid and inaccurate measure of HR.

6.1 Introduction

The spatial and temporal confidence frameworks improved HR estimation and presented a confidence measurement that enables rejecting incorrect values that are invalid and do not match the actual HR. The confidence framework involves a set of parameters along a confidence metric that select and assess the HR estimate. Both spatial and temporal methods were evaluated through comparison with the base VM. The methods provide a set of measures including the location of peaks, the peak magnitudes, and the FFT_Ratio. The measures are parametrized as features which will be used in this analysis to build decisions around the VM results. Machine Learning is used for data analysis to determine trends based on the response of features from both spatial and temporal frameworks methods to ensure an overall confidence framework which will increase the reliability in predicting the HR estimate. The aim is to identify the confidence assessment in the predicted HR and reject the inaccurate and incorrect values.
6.2 Methods

This section includes the major steps for the application of Machine Learning on all the features extracted from the STDA and TFA methods, as well as further assessment of performance through comparison with the base VM method. Fig. 6.1 shows the generic framework in which the confidence parameters from both STDA and TFA are assigned to specific features that will be combined into distinct pools of features such as measures of peak location, magnitude, and FFT_Ratio. The feature dataset for each method is individually imported into Machine Learning but will eventually fuse altogether as part of the overall framework.

All the data from the assigned features (features dataset) are imported into specific Machine Learning models and the test results guide the feature set reduction process using the accuracy score. The Machine Learning model results assess which combination of features provides the best HR prediction assessment. The selection of the features is crucial to filtering those with low value in classification of HR. The low value is determined by a decreased level of accuracy test scoring that indicates the number or percentage of times the Machine Learning model made the correct prediction.

The full implementation runs on MATLAB through the “Classification Learner” app, and the models run on the features dataset are resampled through five cross validations, which are Fine Tree, Linear Discriminant, Logistic Regression, Gaussian Naïve Bayes, Linear SVM, KNN, SVM Kernel, Boosted Trees, and Neural Network.
6.2.1 Feature Assignment for Spatial Time Domain Averaging (STDA) Method

The STDA method in the Chapter 4 confidence parameters uses the first and second largest peaks from the FFT plots results as well as the FFT_Ratio, which are assigned unique names to prepare a feature set for the STDA method as shown in Fig. 6.1. Each of the five parameters are assigned feature names that resemble the parameter (Table 6.1). The feature dataset is prepared through a table (Fig. 6.2) with five entries assigned for each of the 19 subjects pertaining to the inner five time windows common with the TFA method. Fig. 6.2 shows an overview of the table elements showing the five columns for the five features for a given subject.
Table 6.1. Feature name and description for STDA method

<table>
<thead>
<tr>
<th>Feature Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>STDAH1</td>
<td>Largest Peak Indicated HR (HR_{LP})</td>
</tr>
<tr>
<td>STDAN1</td>
<td>Normalized Magnitude for (HR_{LP}) (NM_{LP})</td>
</tr>
<tr>
<td>STDAH2</td>
<td>Second Largest Peak Indicated HR (HR_{SLP})</td>
</tr>
<tr>
<td>STDAN2</td>
<td>Normalized Magnitude for (HR_{SLP}) (NM_{SLP})</td>
</tr>
<tr>
<td>FR</td>
<td>FFT Ratio between (NM_{LP}) and (NM_{SLP})</td>
</tr>
</tbody>
</table>

Fig. 6.2. Overview of the prepared table for Machine Learning with all the features in Table 6.1 for STDA method.

6.2.2 Feature Assignment for Temporal FFT Averaging (TFA) Methods

The TFA method in Chapter 5 identifies parameters relating to the same measures as for STDA method. Table 6.2 shows all the features for the three TFA cases and Fig.
6.3 resembles the table of features for the TFA method. Each subject is assigned five entries for the five time windows, and each time window has five different features for each case. The full feature table is then imported into Machine Learning by applying specific learning and training models.

Table 6.2. Feature name and description for TFA method.

<table>
<thead>
<tr>
<th>Feature Name</th>
<th>Confidence Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>H14</td>
<td>$HR_{LP}$ for TFA case 1</td>
</tr>
<tr>
<td>H14</td>
<td>$HR_{LP}$ for TFA case 1</td>
</tr>
<tr>
<td>H19</td>
<td>$HR_{LP}$ for TFA case 2</td>
</tr>
<tr>
<td>H114</td>
<td>$HR_{LP}$ for TFA case 3</td>
</tr>
<tr>
<td>N14</td>
<td>$NM_{LP}$ for TFA case 1</td>
</tr>
<tr>
<td>N19</td>
<td>$NM_{LP}$ for TFA case 2</td>
</tr>
<tr>
<td>N114</td>
<td>$NM_{LP}$ for TFA case 3</td>
</tr>
<tr>
<td>H24</td>
<td>$HR_{SLP}$ for TFA case 1</td>
</tr>
<tr>
<td>H29</td>
<td>$HR_{SLP}$ for TFA case 2</td>
</tr>
<tr>
<td>H214</td>
<td>$HR_{SLP}$ for TFA case 3</td>
</tr>
<tr>
<td>N24</td>
<td>$NM_{SLP}$ for TFA case 1</td>
</tr>
<tr>
<td>N29</td>
<td>$NM_{SLP}$ for TFA case 2</td>
</tr>
<tr>
<td>N214</td>
<td>$NM_{SLP}$ for TFA case 3</td>
</tr>
<tr>
<td>FR4</td>
<td>FFT_Ratio between N14 and N24</td>
</tr>
<tr>
<td>FR9</td>
<td>FFT_Ratio between N19 and N29</td>
</tr>
<tr>
<td>FR14</td>
<td>FFT_Ratio between N114 and N214</td>
</tr>
</tbody>
</table>
6.2.3 Features Assignment for Base VM

The base VM feature assignment is prepared for Machine Learning as the results will help evaluate the level of efficiency of the STDA and TFA methods and test whether integrating base VM in addition to both methods provides an even better performance or results in a degraded performance. The feature set for the base VM is common to STDA and TFA, having the same dataset with same features for all inner five time windows for a given subject.

Table 6.3. Feature name and description of confidence measure for base VM method.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Confidence Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>VMH1</td>
<td>$HR_{LP}$</td>
</tr>
<tr>
<td>VMN1</td>
<td>$NM_{LP}$</td>
</tr>
<tr>
<td>VMH2</td>
<td>$HR_{SLP}$</td>
</tr>
<tr>
<td>VMN2</td>
<td>$NM_{SLP}$</td>
</tr>
<tr>
<td>VMFR</td>
<td>FFT_Ratio</td>
</tr>
</tbody>
</table>
6.2.4 Application of Machine Learning and Features Selection

After preparing the full feature set, training models are run on the feature dataset for STDA, TFA, and base VM one at a time. First, the training models run on the full feature set, then several features are eliminated, leaving only the most relevant features. The elimination of features is on basis of the model’s test results via the test score accuracy.

6.2.4.1 Feature Selection

The aim of the feature selection is to reduce the dimensionality to help understand the significance and the impact of each confidence measure/parameter to correctly predict and assess a chosen HR. Since the full feature dataset is <100 data points, scoring and ranking models are not highly effective in this case. The feature reduction approach is
through correlation analysis that eliminates features. An additional step in the feature preparation is assignment of the features from the three methods into distinct pools (Fig. 6.5). For example, all the largest peaks indicating HR ($HR_{LP}$) from the three methods are assigned to a specific pool and labelled “H1,” and all the rest are assigned and marked accordingly.

![Fig. 6.5. Assigning all features from STDA, TFA, and base VM across distinct pools based on the confidence measure.](image)

Fig. 6.5 summarizes the full feature dataset table that combines the two methods showing the 95 entries pertaining to the results for the inner five time windows. The additional column for base VM is not shown in the table; however, it is also part of the whole table, having been removed in the given table for clarity purposes.
6.2.4.2 Importing Feature Dataset into Training Models

The Feature Datasets from all three methods are imported into Machine Learning for training and testing. First, all the features are used, followed by feature reduction by eliminating specific feature pools each time. Initially, only STDA feature dataset is analyzed through a set of training models, followed by TFA, and then the base VM features, each imported independently. Then, all three methods are combined into one larger dataset with the same features collection.

6.3 Results

The results section is divided into three sections which present a summary of the training model’s scoring using accuracy score for the three distinct methods (STDA, TFA, and base VM) as well as after fusing them altogether.
6.3.1 Results for Base VM Method

Table 6.4 shows the accuracy results (in %) for a list of training models on the base VM feature dataset. The table reports a highest accuracy value of 83.2% after reducing the feature dataset by eliminating feature pools H1, H2, and F.

<table>
<thead>
<tr>
<th>Machine Learning Model</th>
<th>Features Set</th>
<th>All</th>
<th>N1, N2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fine Tree</td>
<td>74.7</td>
<td>77.9</td>
<td></td>
</tr>
<tr>
<td>Linear Discriminant</td>
<td>77.9</td>
<td>78.9</td>
<td></td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>75.8</td>
<td>78.9</td>
<td></td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>77.9</td>
<td>80.0</td>
<td></td>
</tr>
<tr>
<td>Linear SVM</td>
<td>80.0</td>
<td>74.7</td>
<td></td>
</tr>
<tr>
<td>Fine KNN</td>
<td>81.1</td>
<td>83.2</td>
<td></td>
</tr>
<tr>
<td>SVM Kernel</td>
<td>78.9</td>
<td>75.8</td>
<td></td>
</tr>
<tr>
<td>Boosted Trees</td>
<td>74.7</td>
<td>78.9</td>
<td></td>
</tr>
<tr>
<td>Neural Network</td>
<td>75.8</td>
<td>81.1</td>
<td></td>
</tr>
</tbody>
</table>

6.3.2 Results for STDA Method

Table 6.5 shows the accuracy score for the training models on the STDA feature dataset. The results indicate similar accuracy scores before and after features reduction; however, the highest accuracy score comes from reducing down to N1 and N2 only. The Fine Tree with only two features is an indication of the effectiveness of STDA to use only two sets of measures for N1 and N2 to correctly predict HR. The confusion matrix for the Fine Tree model in Fig. 6.7 shows that that model performs best in predicting the true HR and rejecting invalid ones. However, there is a tie in the observations when VM-incorrectly predicts HR.
Table 6.5. Accuracy Score for training models for STDA method.

<table>
<thead>
<tr>
<th>Machine Learning Model</th>
<th>Features Set</th>
<th>All</th>
<th>N1, N2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fine Tree</td>
<td></td>
<td>80</td>
<td>82.1</td>
</tr>
<tr>
<td>Linear Discriminant</td>
<td></td>
<td>80</td>
<td>74.7</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td></td>
<td>78.9</td>
<td>75.8</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td></td>
<td>81.1</td>
<td>76.8</td>
</tr>
<tr>
<td>Linear SVM</td>
<td></td>
<td>78.8</td>
<td>74.7</td>
</tr>
<tr>
<td>Fine KNN</td>
<td></td>
<td>74.7</td>
<td>69.5</td>
</tr>
<tr>
<td>SVM Kernel</td>
<td></td>
<td>81.1</td>
<td>74.7</td>
</tr>
<tr>
<td>Boosted Trees</td>
<td></td>
<td>74.7</td>
<td>71.6</td>
</tr>
<tr>
<td>Neural Network</td>
<td></td>
<td>80.0</td>
<td>72.8</td>
</tr>
</tbody>
</table>

![Confusion Matrix](image)

Fig. 6.7. Confusion Matrix for observatory performance of the Fine Tree model on STDA feature dataset.
6.3.3 Results for TFA Method

Table 6.6 shows the accuracy score for the training models on the feature dataset for the three cases of TFA. The table reports similarities in the accuracy score; however, like in the STDA case, reducing the feature set did indicate improved accuracy, giving the highest accuracy score of 89.5% when the pool reduced to only H1. This indicates that TFA is a very efficient tool that improves HR measurement and therefore indicates proper performance for the TFA method in general.

Table 6.6. Accuracy Score for training models on selected feature sets for all three cases of TFA method.

<table>
<thead>
<tr>
<th>Machine Learning Model</th>
<th>Features Set</th>
<th>All</th>
<th>H1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fine Tree</td>
<td>83.2</td>
<td>83.2</td>
<td></td>
</tr>
<tr>
<td>Linear Disc.</td>
<td>83.2</td>
<td>87.4</td>
<td></td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>78.9</td>
<td>87.4</td>
<td></td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>77.0</td>
<td>69.5</td>
<td></td>
</tr>
<tr>
<td>Linear SVM</td>
<td>88.4</td>
<td>89.5</td>
<td></td>
</tr>
<tr>
<td>KNN</td>
<td>86.3</td>
<td>88.4</td>
<td></td>
</tr>
<tr>
<td>SVM kernel</td>
<td>76.9</td>
<td>85.3</td>
<td></td>
</tr>
<tr>
<td>Boosted Trees</td>
<td>82.1</td>
<td>74.7</td>
<td></td>
</tr>
<tr>
<td>Neural Network</td>
<td>76.8</td>
<td>89.5</td>
<td></td>
</tr>
</tbody>
</table>

6.3.4 Results from Combining STDA and TFA Features Dataset

Table 6.7 shows the accuracy score for the training models on the combined feature dataset merging both the STDA and TFA methods. Adding the base VM features dataset reduced the accuracy scores, and, in fact, STDA and TFA together show an increase in
the accuracy scores compared to using each method individually in Tables 6.5 and 6.6. The combined result is a 92.6% accuracy performance, which is the highest score so far, indicating that the best performance in correctly predicting the HR and rejecting inaccurate values comes from having a reduced feature set.

Table 6.7. Accuracy Score for training models for features sets for combined STDA and TFA.

<table>
<thead>
<tr>
<th>Accuracy Score (in %)</th>
<th>Features Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine Learning Model</td>
<td>All</td>
</tr>
<tr>
<td>Fine Tree</td>
<td>84.2</td>
</tr>
<tr>
<td>Linear Disc.</td>
<td>84.2</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>70.5</td>
</tr>
<tr>
<td>Gaussian Naïve Bayes</td>
<td>69.5</td>
</tr>
<tr>
<td>Linear SVM</td>
<td>90.5</td>
</tr>
<tr>
<td>KNN</td>
<td>83.2</td>
</tr>
<tr>
<td>SVM kernel</td>
<td>84.2</td>
</tr>
<tr>
<td>Boosted Trees</td>
<td>74.7</td>
</tr>
<tr>
<td>Neural Network</td>
<td>81.1</td>
</tr>
</tbody>
</table>

The confusion matrices in Fig. 6.8-6.14 summarize the observations based on the models resulting in the highest accuracy score across several combinations of feature sets. The aim is to understand the variation in the observations based on a specific confidence measure. From the matrices, it appears that Fine Tree and SVM kernel provide the best performance, as they show the highest accuracy scores where all the correct HR are classified as correct HR and having none of them misclassified. Generally, models perform well in identifying a correct HR (True Class =1) and classifying it as correct (Predicted =1). However, it appears that when the HR is incorrect (True Class =0), the
observations are close to a ~50% scoring between predicted correct or incorrect. There is an imbalance between the positive (correct HR) and negative (incorrect HR) classes.

Fig. 6.8. Confusion matrix for Linear SVM with 90% accuracy score for STDA and TFA feature set
Fig. 6.9. Confusion matrix with 92.6% accuracy score using SVM kernel for STDA and TFA feature set

Fig. 6.10. Confusion matrix for 90% accuracy score using Fine Tree model for STDA and TFA feature set
Fig. 6.11. Confusion matrix for 78.9% accuracy test score using Linear SVM model for STDA and TFA feature set.

Fig. 6.12. Confusion matrix for 82.1% accuracy test score using SVM Kernel model for STDA and TFA feature set.
Fig. 6.13. Confusion matrix for 89.5% accuracy test score using SVM Kernel model for STDA and TFA feature set.

Fig. 6.14. Confusion matrix for 82.1% accuracy test score using Fine Tree model for STDA and TFA feature set.
6.4 Discussion

The high accuracy scoring for the Machine Learning models is an essential indicator for evaluating the models to properly classify the correct and incorrect VM-predicted HR measures. The results show that the high accuracy scoring for classification of HR seen among all three methods (base VM, TFA, and STDA) indicates efficiency and essentiality of the chosen confidence measures. This indicates that VM results are essential measures for HR estimation. The results of the Machine Learning indicate that STDA and TFA are successful in improving HR measurement, as they provide improved accuracy scores alternate to the base VM method. This indicates that both methods are crucial algorithms of VM and provide better measures of confidence parameters compared to VM alone. Nevertheless, accuracy scores from combined datasets (TFA and STDA) demonstrates the cruciality of both methods, as combining them increases the reliability in precisely identifying correct HR from invalid measures. The Machine Learning implementation can improve HR estimation by rejecting incorrect HR to a certain extent.

There was very high performance when identifying the correct HR; however, for incorrect HR, performance is a bit problematic. There seems to be a class imbalance issue that warrants further exploration in several areas, such as the pre-processing stages before Machine Learning modelling and training. Perhaps exploring the population closely can help in the resampling and re-ordering of the data for more objective learning and training. Nevertheless, Machine Learning can also improve the confidence framework for estimating HR by optimizing performance with a reduced feature set. Lastly, Machine Learning is a very essential component in VM at this stage and can be used in the future
when considering a larger population, such as using a larger number of subjects. Since the dataset was collected from subjects that were selected during the COVID-19 pandemic, accessibility to a larger variety of sample data points was restricted. Finally, Machine Learning will be powerful, as TFA has shown issues in complexity in several cases and observing the data can be tedious. Machine Learning uses programmed algorithmic models that will remove issues with human error that will help increase reliability of VM in terms of time, cost, and reducing feature sets in ways unlikely to happen solely by human judgement.
Chapter 7: Conclusion

Heart Rate (HR) is an essential health parameter that is used for maintaining and assessing the overall wellbeing of individuals. The recent pandemic and the inclined focus of research towards senior assisted living technology paved the way to incorporate highly advanced vital-sign-assessment tools such as Video Magnification (VM) into measurement. VM is a non-contact and remote method for HR measurement that works by examining the minuscule changes in color of the skin in the video caused by blood flow due to systole and diastole processes. However, previous work highlighted and explored the limitations that hinder the performance of VM. Limitations like skin color, body motion, and illumination can cause changes in skin color which reveal non-true variations of color changes in video that do not match those correlating to HR. Such limitations can cause interferences with the VM results, preventing identification of those correlated with HR. Previous work explored different disciplinary aspects in VM to limit the impact of specific types of limitations; however, this thesis establishes a confidence framework that will be used in the estimation of HR. The framework supports the selection and identification of the VM results that match HR and reject invalid results which may be challenging due to the limitations.

The first part of the confidence framework explored the spatial domain (Chapter 4) aspect of the VM algorithm through combining ROIs using two distinct methods. The first method, the time domain method, occurs at the early stages of VM and averages all the pixel elements for all the selected ROIs across the video frames (STDA method). The second method, frequency domain averaging (SFDA), involves averaging all the frequency results from VM. The methods were tested on the 19 NRC subjects and the
confidence framework through both methods includes analysis of the spectral peaks for
the FFT results to help identify the correct HR estimate. The SFDA methods showed
correct HR as the highest spectral peak for 16 subjects, and STDA showed the correct
HR for 18 subjects. The average confidence metric was higher for the STDA method (2.9)
than the SFDA method (2.3) or a single forehead ROI case (2.3), and the confidence
metric was generally lowest (<2) for subjects with more body motion or darker skin tones
that challenge VM.

The confidence framework explored VM in the temporal domain (Chapter 5)
looking at VM results from consecutive and overlapping time windows, involving the
time domain and frequency domain aspects of VM. This technique is the temporal FFT
averaging method (TFA), which combines the preceding, current, and subsequent FFT
results for those time windows and chooses the largest frequency as new frequency. The
confidence framework analyzes the spectral peaks for the new FFT to be able to reliably
select the correct HR. The temporal averaging technique improved the performance of
VM and improved overall HR accuracy from 61.1% to 68.4%. HR errors were shown to
be associated with motion artifacts and/or darker skin tones. The confidence metric is
generally proportional to the accuracy of the HR estimate, providing an indication of
when VM-based HR measurements are reliable.

Finally, Chapter 6 merges both STDA and TFA to assess the VM predicted HR
from both methods based on accuracy levels through the implementation of Machine
Learning models. The high accuracy test scores validated the improvements and the
efficiency of both the spatial and temporal confidence frameworks to improve the HR
prediction and estimation.
Future work can also explore different responses of the VM predictions through confidence measures caused by variations in illumination and variations in body motions. This leads to a parametrized set of measures as part of the overall confidence framework.
Appendices

Appendix A

A.1 Mapping of NRC Subjects

This section of the appendix shows the new mapping of the NRC subjects by a former research fellow as part of an earlier research work and after excluding Subjects 1, 5, and 18 (in red). The mapping includes reordering of subjects and mapping all the video file naming accordingly.

<table>
<thead>
<tr>
<th>Before Exclusion</th>
<th>After Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>S1</td>
</tr>
<tr>
<td>S2</td>
<td>S1</td>
</tr>
<tr>
<td>S3</td>
<td>S2</td>
</tr>
<tr>
<td>S4</td>
<td>S3</td>
</tr>
<tr>
<td>S5</td>
<td>S4</td>
</tr>
<tr>
<td>S6</td>
<td>S5</td>
</tr>
<tr>
<td>S7</td>
<td>S5</td>
</tr>
<tr>
<td>S8</td>
<td>S6</td>
</tr>
<tr>
<td>S9</td>
<td>S7</td>
</tr>
<tr>
<td>S10</td>
<td>S8</td>
</tr>
<tr>
<td>S11</td>
<td>S9</td>
</tr>
<tr>
<td>S12</td>
<td>S10</td>
</tr>
<tr>
<td>S13</td>
<td>S11</td>
</tr>
<tr>
<td>S14</td>
<td>S12</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>S15</td>
<td>S13</td>
</tr>
<tr>
<td>S16</td>
<td>S14</td>
</tr>
<tr>
<td>S17</td>
<td>S15</td>
</tr>
<tr>
<td>S18</td>
<td></td>
</tr>
<tr>
<td>S19</td>
<td>S16</td>
</tr>
<tr>
<td>S20</td>
<td>S17</td>
</tr>
<tr>
<td>S21</td>
<td>S18</td>
</tr>
<tr>
<td>S22</td>
<td>S19</td>
</tr>
</tbody>
</table>
Appendix B

B.1 Code Stack for ROI Selection and Acquisition

```matlab
roi=1; % Only one ROI is selected but can increase to multiple ROIs depending on application
ry=cell(1,roi); %
rx=cell(1,roi);

for q=1:1:roi % loop once as only 1 ROIs is used here
    ext=readv(floor((FrameNum/2)+200)); % select a specific single frame
    [x,y,p]=impixel(ext) % represent the whole frame image into a grid
    showing all coordinates for selecting specific locations
    xx=x';
    yy=y';
    % pixel columns and row
    col=[min(xx):max(xx)]; % obtain all the pixels in the horizontal region and
    store into a single vector
    row=[min(yy):max(yy)];% obtain all the vertical region store
    all the pixels and store into a single pixel
    cx(q)=col; % A single cell stores all the columns pixels
    ry(q)=row; %A single cell stores all the row pixels

end
```
References


[13] “Role of Elevated Heart Rate in the Development of Cardiovascular Disease in Hypertension | Hypertension.”


