Computational Approaches for Monitoring of Health Parameters and Their Evaluation for Application in Clinical Setting.

Mohammad Adibuzzaman
Marquette University

COMPUTATIONAL APPROACHES FOR MONITORING OF HEALTH PARAMETERS AND THEIR EVALUATION FOR APPLICATION IN CLINICAL SETTING

by
Mohammad Adibuzzaman

A Dissertation submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy

Milwaukee, Wisconsin
May, 2015

ABSTRACT
COMPUTATIONAL APPROACHES FOR
MONITORING OF HEALTH PARAMETERS
AND
THEIR EVALUATION FOR
APPLICATION IN CLINICAL SETTING

Mohammad Adibuzzaman, B.S., M.S.

Marquette University, 2015

The algorithms and mathematical methods developed in this work focus on using computational approaches for low cost solution of health care problems for better patient outcome. Furthermore, evaluation of those approaches for clinical application considering the risk and benefit in a clinical setting is studied. Those risks and benefits are discussed in terms of sensitivity, specificity and area under the receiver operating characteristics curve. With a rising cost of health care and increasing number of aging population, there is a need for innovative and low cost solutions for health care problems. In this work, algorithms, mathematical techniques for the solutions of the problems related to physiological parameter monitoring have been explored and their evaluation approaches for application in a clinical setting have been studied. The physiological parameters include affective state, pain level, heart rate, oxygen saturation, hemoglobin level and blood pressure. For the mathematical basis development for different data intensive problems, eigenvalue based methods along with others have been used in designing innovative solutions for health care problems, developing new algorithms for smart monitoring of patients; from home monitoring to combat casualty situations. Eigenvalue based methods already have wide applications in many areas such as analysis of stability in control systems, search algorithms (Google Page Rank), Eigenface methods for face recognition, principal component analysis for data compression and pattern recognition. Here, the research work in 1) multi-parameter monitoring of affective state, 2) creating a smart phone based pain detection tool from facial images, 3) early detection of hemorrhage from arterial blood pressure data, 4) noninvasive measurement of physiological signals including hemoglobin level and 5) evaluation of the results for clinical application are presented.

ACKNOWLEDGMENTS

Mohammad Adibuzzaman, B.S., M.S.

It is my great pleasure to express my gratitude and thank my supervisors Dr. Sheikh Iqbal Ahamed and Dr. Stephen Merrill for their continuous support, encouragement, and guidance for my work.

I also thank Dr Dennis Brylow, Dr Praveen Madiraju, and Dr Bhagwant Sindhu for taking their valuable time for serving as committee members.

I am grateful to Dr. Richard Love for his enthusiasm, innovative ideas to solve health care problems. His critical thinking about the methodologies helped me keep focus on my research work. His passion about helping the underprivileged and goal for 'One Hoss-Shay Life' for the people encouraged me to work in healthcare.

I also want to thank Dr. David Strauss. Under his leadership at the Food and Drug Administration, I learnt many things about how to do research and approach a problem scientifically. His focus on details and questions regarding the validity of the approach helped me keep focus and lead me to the next steps. I believe these helped me significantly to be a better researcher. I would like to express my gratitude to Dr. Christopher Scully and Dr. Loriano Galeotti for their mentor-ship, helping me acquire the domain knowledge and persistence for moving the work forward.

Special thanks to Dr. Richard Povinelli for his guidance with deep understanding of the the machine learning and signal processing techniques.

I would also take the opportunity to thank Dr. Joshua Field, Dr. Stephen Hargarten, Dr. Munirul Haque for their support and expertise in the respective areas.

And last but not the least, I thank my family, my parents and brothers, Ferdaus Kawsar and Nafees Ahmed for their unconditional love and support for me. I also like to thank my friends GM Tanim, Prachi Pradeep, Imran Reza, Jahangir Mazumdar for supporting me on my journey in every possible ways.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

Assessment of different health parameters including pain level, physiological parameters such as blood pressure, heart rate, and hemoglobin level is important for multiple medical conditions. Traditional monitoring systems, if they exist, for these health parameters require sophisticated devices and complex systems which makes the application of these systems difficult for many situations such as remote monitoring. The application of these systems is also difficult in developing countries where there is significant lack of laboratory and transportation facilities. This situation also persists in developed countries where the rising health care cost and increasing number of aging population demands affordable solutions for monitoring patients with a constant feedback loop involving doctors and patients[43]. The widespread availability of mobile phones both in developed and developing countries makes the computational power and communication facilities available for such systems. However, there is significant lack of research for bridging the gap between the demand for systems that would address this problem. This dissertation attempts to bridge the gap by building such systems by analyzing the different issues related to deploying such systems and developing algorithms for accurate detection and prediction of the health parameters. This dissertation focuses on eigenvalue based methods for algorithm development and performance evaluation in three different applications. These applications address the challenges related to 1) algorithm and system development for monitoring health parameters including affective states, pain level and vital signs, 2) finding effective ways to interpret the health parameters in clinical setting for intervention by the clinicians 3) creating a closed loop feedback system involving the patients,clinicians and health care personnel and 4) evaluating such systems to be applicable in a clinical

setting.

1.1 Dissertation focus

This study focused on the application of low cost solutions for health parameter monitoring systems. Among these solutions are developing an eigenvalue based method for smart phone based pain monitoring system from facial expression, a new vital sign development for detection of hemorrhage from blood pressure waveform, detecting hemoglobin level from the video image of the camera of a smart phone and evaluating algorithms for application in a clinical setting.

1.2 Dissertation organization

Chapter 2 of this dissertation describes the different mathematical tools and the foundation of the algorithm development. This chapter lays a common mathematical framework based on eigenvalues including the geometric interpretation of the eigenvalues, their role in principal component analysis, and mixing rate of the Markov chain. It also focuses on the health parameters: emotion, pain level, blood pressure, heart rate and hemoglobin level and the difficulties in assessing those parameters in terms of application and algorithm. Chapter 3 and 4 presents the eigenvalue based eigenface method for multimodal affect detection and pain level detection. Chapter 3 describes how the eigenvalue based method is used for multimodal affect detection from facial image and energy expenditure with Naïive Bayes method for the fusion. Chapter 4 describes the eigenvalue based method used for pain assessment and then we described the techniques and hypothesis for improving the model accuracy. In chapter 5, a mathematical method development is presented for eigenvalues of the transition probability of a Markov chain created using time series data of arterial blood pressure and found important relationship between second largest eigenvalue and blood loss or hemorrhage. Chapter 6 and 7 are dedicated for model development for vital sign monitoring using finger tip video images. In these chapters we also describe

the different mathematical models and initial results for detecting heart rate, perfusion index, oxygen saturation and hemoglobin level. Chapter 8 identifies the issues related to algorithm evaluation for clinical use. For each of this chapter, we describe the motivation and importance of the problem, mathematical methods, results and evaluation.

1.3 Publications

- **Mohammad Adibuzzaman**, Colin Ostberg et al. ***Assessment of Pain Using a Smart Phone*** in Proceedings of COMPSAC 2015.

- **Mohammad Adibuzzaman**, George C. Kramer, Loriano Galeotti et al. ***The Mixing Rate of the Arterial Blood Pressure Waveform Markov Chain is Correlated with Shock Index During Hemorrhage in Anesthetized Swine*** in Proceedings of EMBC 2014, Chicago, USA

- **Mohammad Adibuzzaman**, Loriano Galeotti et al. ***A Novel Index to Monitor Physiological Systems from the Arterial Blood Pressure Waveform during Hemorrhage*** in MCMi Regulatory Science Symposium, 2014, Maryland, USA

- **Mohammad Adibuzzaman**, Sheikh Ahamed et al. ***A Personalized Model for Monitoring Vital Signs using Camera of the Smart Phones*** in Proceedings of SAC 2014, Seoul, Korea

- **Mohammad Adibuzzaman**, Niharika Jain,Nicholas Steinhafel, Munir Haque, Ferdaus Kawsar, Richard Love, Sheikh Iqbal Ahamed. ***Towards In Situ Affect Detection in Mobile Devices: A multimodal Approach*** in proceedings of Research in Adaptive and Convergent Systems(RACS'13) (Best Paper Award)

- Md. Munirul Haque, **Md Adibuzzaman** et al. ***IRENE: Context Aware Mood Sharing for Social Network*** in KASTLES 2011 Workshop.

- **Mohammad Adibuzzaman**, Loriano Galeotti, Richard A. Gray, George Kramer, David G. Strauss, Stephen Merrill. ***Markov chain methods in identifying early acute hypotensive episodes*** in MCMi Regulatory Science Symposium, 2013

- **Mohammad Adibuzzaman**, Loriano Galeotti, Richard A. Gray, George Kramer, David G. Strauss, Stephen Merrill. ***Novel physiological monitoring strategies for early detection of hemorrhage*** in Student Poster Session 2013 at the U.S. Food and Drug Administration.

- **Mohammad Adibuzzaman**, Christopher Scully, Loriano Galeotti, David Strauss, Stephen Merrill. **Evaluation of Machine Learning Algorithms for Multi-Parameter Patient Monitoring** in Student Poster Session 2014 at the U.S. Food and Drug Administration.

- Under review

  - **Mohammad Adibuzzaman**, Sheikh Iqbal Ahamed, Richard Love. ***Assessment of Hemoglobin Level using Mini-video image*** in preparation.

Chapter 2

Background

This chapter lays the outline for the background on the mathematics of the algorithms that has been used in the application areas for the remaining chapters. Here the basics of eigenvalues are described, which has applications in chapter 3, 4 and 5. It has been used for extracting features from face images using eigenfaces in chapter 3 and chapter 4; and in using the second largest eigenvalue(mixing rate) of the transition probability matrix of Markov chain in chapter 5.

2.1 Eigenvalues

Definition

If a nonzero vector (eigenvector of a matrix) is multiplied by a matrix, the eigenvalue of a matrix is the constant by which amount the vector is scaled up or scaled down. Eigenvalue is defined as

$$B \times v = \lambda \times v$$

where v is the right eigenvector of the matrix B and λ is the eigenvalue.

Eigenvalues have an interesting property: when multiplied by the eigenvalue, if there is an eigenvector, the result is not rotated. They are just scaled up or scaled down if multiplied by an eigenvalue which is a real number.

This idea can be extended for any vector since any vector can be represented by a basis whose reference vectors are eigenvectors of the transition matrix in the case there is a complete set of eigenvectors and the eigenvalues are distinct.

$$B^i \times v = B^i \times v_1 + B^i \times v_2 = {\lambda_1}^i \times v_1 + {\lambda_2}^i \times v_2$$

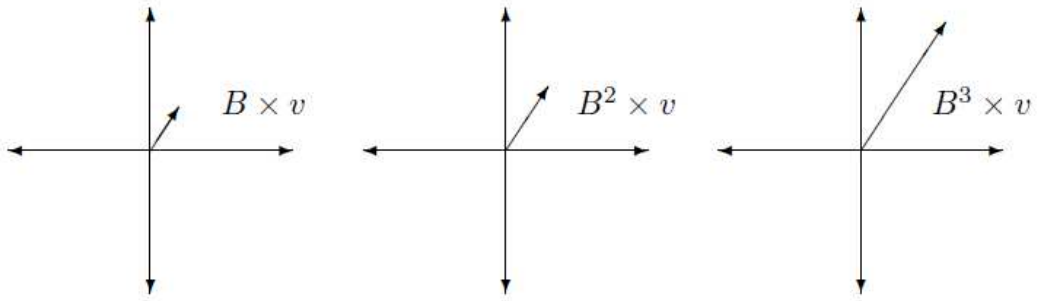


Figure 2.1: Repeated multiplication of the eigenvector v with the matrix B scales up the eigenvector. Here the eigenvalue λ is assumed to be greater than 1.

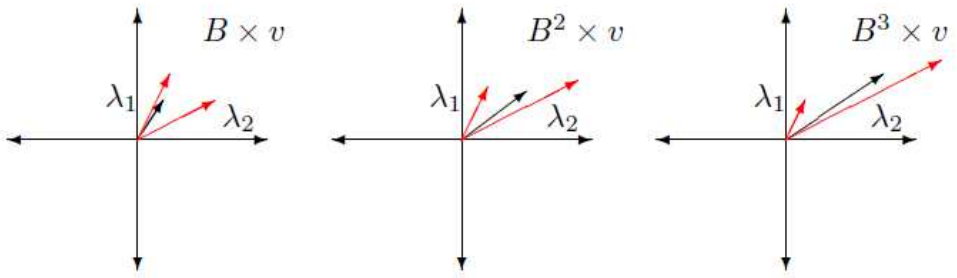


Figure 2.2: Repeated multiplication of the eigenvector v with the matrix B with more than one eigenvalue. Here the eigenvalue $\lambda_2 > \lambda_1$, and as a result, the resultant vector aligns with the eigenvector v_2, corresponding to eigenvalue λ_2.

As a result, for any iterative methods which involve multiplication of a vector by a matrix in a repetitive form, the component related to the maximum eigenvalue, if there is one, dominates in the long run.

- This is the basis for the algorithms used for smart monitoring of health parameters.
- Principal component analysis and mixing rate of Markov chain both are based on this interesting property.

2.2 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) highlights the similarities and dissimilarities in a multidimensional data set. It is used to find the patterns in data

using the concept described earlier. This method is widely used for data compression without much loss of information by finding the vectors corresponding to the maximum eigenvalue. Figure 2.3 shows the eigenvectors of the covariance matrix of a two dimensional data set. The red dashed lines are the eigenvectors. The data are more distributed along with the eigenvector labeled as v_1 for which we have the maximum eigenvalue. As a result, reducing the dimension corresponding to v_2 would not loose much information.

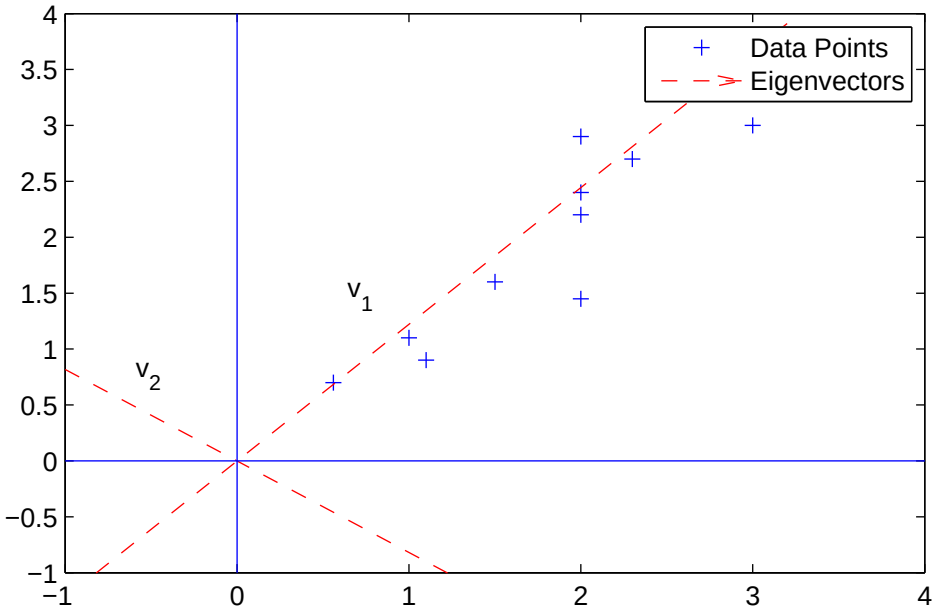


Figure 2.3: The red dashed lines represent the eigenvectors of the covariance matrix. The data points can be represented as a linear combination of the eigenvectors. v_1 is the Principal Component here.

2.3 Eigenvalues and Markov Chain

2.3.1 Markov Process

A discrete Markov chain is a mathematical process that transits from one state to the other and is considered memory less.

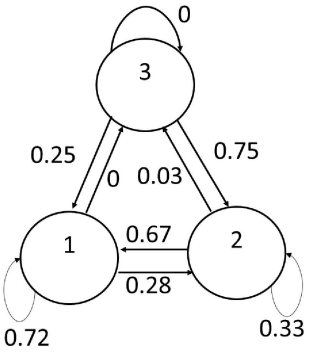


Figure 2.4: A simple Markov chain with three states and their transition probabilities.

$$P = \begin{bmatrix} 0.72 & 0.28 & 0 \\ 0.67 & 0.30 & 0.03 \\ 0.25 & 0.75 & 0 \end{bmatrix}$$

Figure 2.5: The transition probability matrix of the Markov Chain. The row and columns of the matrix represent the transition probability from the state corresponding to the row to the state corresponding to the column. For example, the value of 0.28 at row 1 column 2 indicates that the transition probability of the system from state 1 to state two is 0.28.

2.3.2 Eigenvalues of Markov Chain

Eigenvectors of a Markov chain describes in the long run what would be the probability distribution of the system for being in different states. This uses the same concept as described earlier. If we multiply the transition probability matrix many times with some initial state probabilities, this is the same as multiplying the state probabilities many times with the the eigenvalues. As a result, the eigenvalues that are less than 1 in magnitude, soon make the system converge to the limit distribution. The limit distribution is thus just the left eigenvector corresponding to 1.

- For any transition probability matrix of a Markov chain, there exists an unique eigenvector, π for which the eigenvalue is 1.
- Because all the other eigenvalues are less than one in modulus, after n steps, where n is sufficiently large, the stationary distribution contains only the eigenvector corresponding to the eigenvalue 1.

$$\pi \times P = \pi$$

- For a 2×2 matrix,

$$v \times P^i = v_1 \times P^i + v_2 \times P^i = {\lambda_1}^i \times v_1 + \overbrace{{\lambda_2}^i \times v_2}^{0(\lambda_2<1)}$$

$$= 1^i \times v_1$$

$$= v_1$$

For the previous transition matrix, the normalized eigenvector corresponding to 1 is,

$$\pi = \begin{bmatrix} 0.7014 & 0.2899 & 0.0087 \end{bmatrix}$$

In our case

$$P^{100} = \begin{bmatrix} 0.7014 & 0.7014 & 0.7014 \\ 0.2899 & 0.2899 & 0.2899 \\ 0.0087 & 0.0087 & 0.0087 \end{bmatrix}$$

Each column of the probability distribution becomes the eigenvector normalized.

2.3.3 Mixing Rate of Markov Chain

Mixing Rate or Second Largest Value of Markov Chain

- The largest eigenvalue of a Markov transition matrix is 1.

- The second largest eigenvalue of the matrix determines how fast the chain would converge to the limit distribution.
- The second largest eigenvalue is called the mixing rate of the Markov chain.

Chapter 3

Affect: Smart Phone Based Affect Detection

3.1 Introduction

Affect sensitive applications are being developed in a number of domains which include learning technologies, autism spectrum, gaming, robotics and Human Computer Interaction [48]. Though there has been significant research in the field of affective computing, most of the research is done for unimodal affect detection.

Interestingly, human affective state is never a unimodal expression. Any affective state such as happiness, sorrow, anger etc. almost always involves two or more of the modalities such as facial expression, body movement, speech data and other emotional cues. Thus a multimodal affect detection system may have better accuracy and reliability that has largely been ignored in the literature.

In the field of affective computing, multimodal real time implementation is widely advocated but rarely implemented [48]. Research has been done for affect detection from facial expression, speech data, body gesture, heart rate, skin conductance, pressure sensor and other inputs. Research has been conducted for affect detection using all of these communication channels in a laboratory environment.

In addition to facial expression research work inspired by Ekman [90], there are other approaches such as region based or holistic approach. Instead of different Action Units (AU) proposed by Ekman, region based approach uses certain regions of the face such as eye, eyebrow and mouth. Some of the expression recognition system uses various pattern recognition techniques [72] and geometric and appearance feature-based methods [95].

Beyond facial expression, several affective computing applications focus on detecting affect by using machine learning techniques to identify patterns in physiological activity [96]. These patterns in physiological activity correspond to

different emotions. Previous research was done using facial expression and speech data fusion. In the literature we have found only one work with the fusion of multiple modalities which includes physiological activity [53]. A correlation between automated assessment of mental (or physical health) and the result of the gold-standard surveys with sensor based measurements were found in [81].

There has been much research for multi-modal affect detection in the past decade [96]; but few research studies consider hand-held devices which are equipped with sensors that can be used for affect detection. The affect detection technique for hand-held devices will not require a multitude of the complicated sensors the user has to wear in laboratory environments. Rather, we need a solution that uses the existing sensors of the hand-held devices.

Fortunately, smart phones are equipped with many sensors such as camera, microphone, accelerometer and GPS. Each of these sensors can be used for capturing affective state in different channels like facial expression, speech, activity and context. Furthermore, the power of hand-held and mobile devices is increasing at a tremendous speed. With the advent of mobile technology, many smart phones now have an accelerometer along with a built-in camera. These two sensor devices are critical for automatic affect detection. Accelerometer data could be used for estimation of how much energy a person has exerted over a period of time. Physiopsychology research shows that there is a considerable correlation between energy expenditure and affective states.

Therefore, we choose to use facial expression and physiological activity for multimodal affect detection in natural environment using camera and accelerometer of smart phones using Naïve Bayes fusion. We evaluated the system performance and found significant improvement of our system, which also include energy expenditure, over unimodal system which uses only facial expression.

3.1.1 Contributions

The ability to detect and understand human affective state is at the core of human intelligence which is indispensable from human behaviour. This is also true to understand human decision making process as well as consumer behaviour [42]. In this paper, we present an automated system that can detect affective state in natural everyday setting using sensors available in smart phones. The contribution of this work includes:

- An automated system that can be used in day to day life without any interference to the user.
- This work shows that the system performs better when we do the fusion of facial expression and energy than only from facial expression.

3.2 Related Work

One of the major problems experienced while doing study on affective computing was to correctly identify the human emotions in a laboratory setting [73]. This is because, in contrary to the natural world scenarios, the controlled environment does not offer the natural occurrence of actions and corresponding reactions. Moreover, the laboratory proceedings do not take into account the results of other concurrent activities happening in day-to-day life. In addition to this, the identification of emotions is sometimes based on the reports created by self-analysis and hence suffers from its own set of limitations.

To overcome these factors, a 7-day study was conducted in [44], where data related to the affective states were collected through several means. First source being the ‘in situ’ analysis taking into affect the elimination of bias and independence of axes. Second source of data was the overall ‘end-of-day’ rating and third source accounted the scores given by the third-party raters. The data collected from these

different sources was then triangulated so as to achieve a data set containing mutually agreeing information. This information when fed to a J48 decision tree returned highly accurate results (100% accuracy) in terms of high or low activation states.

For modeling of affective data in natural environment, a study was conducted by [45] so as to collect the sensor data corresponding to physical activity, heart rate and galvanic skin response. The data so obtained was then aggregated and stored in a mobile device which was also used as a user interface. This mobile device was also used to capture audio data related to the subjects involved. The study concluded that the selection of correct time window and having a customized window around annotated events played key roles in obtaining the correct emotion analysis.

The power of mobile devices has been further explored for affect annotation by incorporating the multi-modal technique for assessing physical as well as mental well-being [81]. A study was conducted wherein the subjects were given a device consisting of various sensors which could help collecting data required for the above mentioned assessment. The classification of collected data as speech and activity was done using two-state hidden Markov Models (HMM) and decision-stump classifiers respectively. A correlation between automated assessment of mental, and/or physical health and the result of gold-standard surveys was found so as to stress upon the accuracy of sensor-based measurements.

Healey et al. tries to standardize the affective data annotation in [44] and [45]. But their approach used sensors not only that are available in mobile devices, but also external sensors. Again, a comprehensive study about the classification algorithm was not present. [81] shows the correlation of gold standard surveys with sensor data capture, but that does not provide a study only for mobile device. The participants had to wear other sensor devices for affective data annotation. In our study we overcome both of these problem by using only one smart phone for the user as well as a comparative study of the result about multimodal versus unimodal system. All of these

works provide the techniques and results of using different modalities. But none of the research study uses only smart phones for collecting data. Also, a comparative study between unimodal system versus multimodal system was not present.

3.3 Our Approach

To capture the arousal and valence space, facial expression and energy exertion of a person were used. It is shown in psychology that arousal space can be captured by heart rate, pupil size or energy expenditure; all of them can be captured from the sensors available in smart phones. Here, facial expression was used for valence and energy expenditure was used for the arousal space. Eigenface algorithm was used for detecting facial expression and mean accelerometer data were used for different affective states as the second feature for multi-modal affect detection using Naïve Bayes fusion. In this section, the methods that has been used for the annotation of *in-situ* affective data and the rationale for using facial expression and energy expenditure are described. In the next section, the details of the classifier are described.

3.3.1 Selecting Modalities

Emotion labeling is moderately less work in laboratory environment where the researcher can control the environment, recreate the situation, recording can be done accurately and the person can be interviewed for his/her annotations. However, the emotion labeling can still be flawed since in controlled environment people might act differently, both physiologically and cognitively. *In-situ* capturing of affective state captures the natural data, but it needs more methodical approach for emotion journaling. The participants need to be trained well and the data labeled needs to be verified later. The goal is to minimize the error for establishing the 'ground truth', which defines true affective state for the given data in machine learning algorithm for classification.

According to the Russell's circumplex model of emotion, each affective state can be

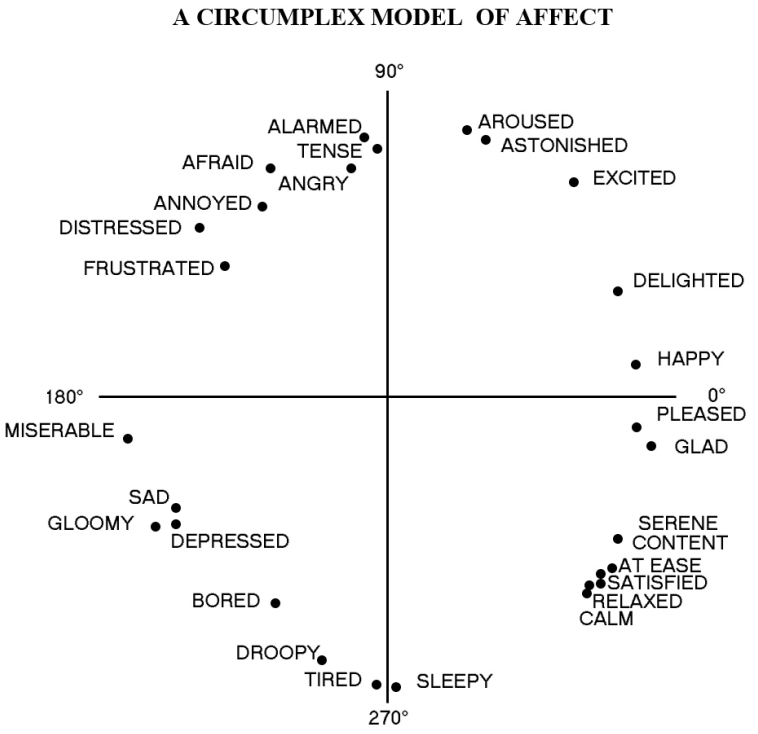


Figure 3.1: Russell's circumplex model of emotion [86].

represented in 2D space [86]. The horizontal axis represents the valence and the vertical axis represents the arousal space. Valence represents how good or bad a person is feeling, and arousal represents how much a person is aroused. Therefore, the hypothesis is that if arousal space data could be captured from accelerometer, the affective states could be classified with better accuracy. For example, for the happy state, this is a positive feeling and a person might have some kind of excitement. On the other hand, for sad feeling, it is a negative feeling and the person may have less movement, which corresponds to less energy expenditure.

3.3.2 Emotion Journaling

Smart phones were used for emotion journaling. Smart phones gives the opportunity for labeling emotion as soon as it occurs with real time sending and

storage capability. Eight participants were recruited for the study, aged between 21-33 all of whom were students. Participants were asked to carry the smart phones and annotate the data for at least 5 times a day for a seven day period. The participants are referred as ”PA” in this work.

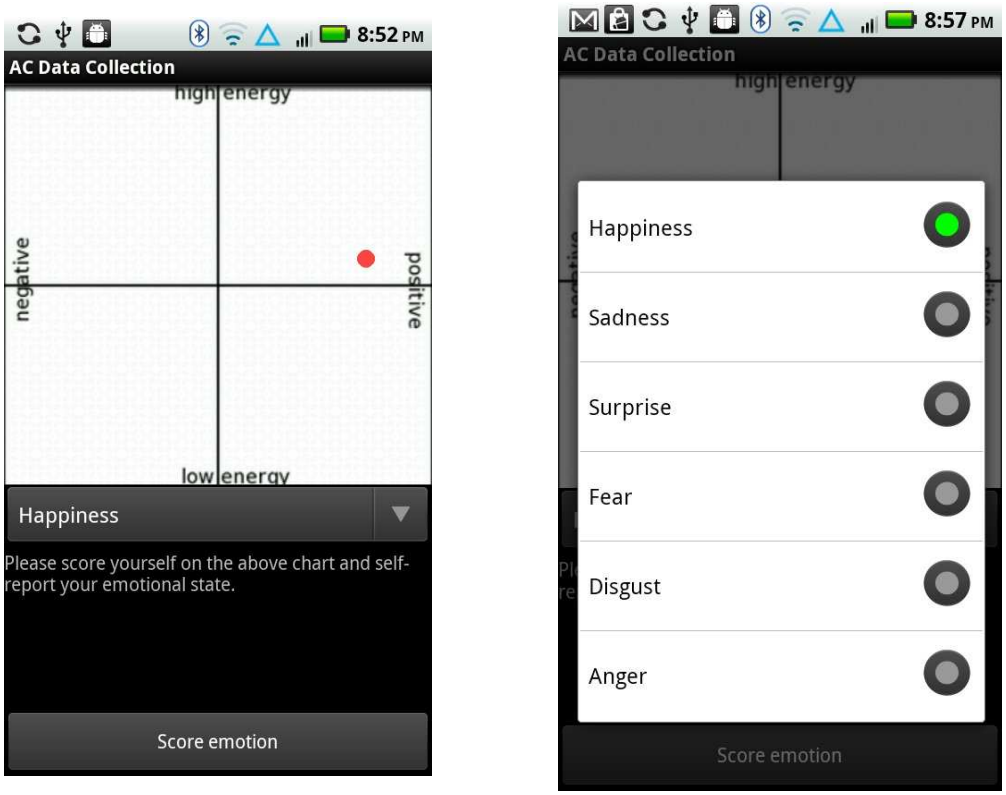


Figure 3.2: Annotation of emotion data: (a) Annotation of affective state using Russells 2D emotional space. (b) Annotation of affective state using radio button.

For the journaling of emotional data, we used camera and accelerometer data of smart phones for facial expression and activity. Also location data was stored using GPS of the smart phones that might give us the context information. Three android phones were used, two Droid X with android operating system of 2.2 and one Samsung Galaxy Nexus with android operating system of 3. PAs were asked to take a facial picture with the smart phones and then label the data. The labeling was done using two sources for capturing the natural feeling. One is using the textitMood-map

which corresponds to Russell's circumplex model and the other is radio button from which the user can pick one from the six basic emotions. Figure 3.2 shows the interface for mood-map as well as the radio buttons. Continuous and fine grained accelerometer data for fifteen minutes before taking the picture and location data were also recorded and then sent to the server using the phones' internet.

3.3.3 Journaling Training

Each participant was asked to carry the smart phone for one week. Before handing over the smart phone to the participants, they were trained on how to use the application for emotional data annotation. During that period, PAs were asked to carry the smart phones six to eight hours a day and label the data whenever any emotional event occurred. They were trained to use the touch based application as well as how to take the picture, use the mood-map and upload the data. Furthermore, there was constant communication between the PAs and the researcher for any question from the participant.

3.3.4 Algorithm Design

The algorithm for affect detection from facial expression and accelerometer data can be discussed in different components; face detection, affective state from facial image, energy expenditure from body movement and fusion using Naïve Bayes. Each of the components are discussed here.

3.3.5 Face Detection

Pixels corresponding to skin are different from other pixels in an image. [72] has shown the clustering of skin pixels in a specific region for skin color modelling in chromatic color space. Though the skin color of persons vary widely based on different ethnicity, research [90] shows that they still form a cluster in the chromatic color space. After taking the image of the subjects, the image was cropped and only the head portion of the image was selected. Then, skin color modeling was used for

extracting the required segment for face from the head image.

3.3.6 Affective State from Facial Image

For this part, a combination of Eigenfaces, Eigeneyes, and Eigenlips methods based on Principal Component Analysis (PCA) [95][96] were used. This analysis method includes only the characteristic features of the face corresponding to a specific facial expression and leaves other features. This strategy reduces the amount of training sample and helps us make our system computationally inexpensive which is one of our prime goals. These resultant images are used as samples for training Eigenface method and M Eigenfaces with highest Eigenvalues were created. The Eigenspace was generated as follows:

- The first step is to obtain a set S with M face images. Each image is transformed into a vector of size N^2 and placed into the set, $S = \gamma_1, \gamma_2, \gamma_3, ..., \gamma_M$
- Second step is to obtain the mean image ψ

$$\psi = \frac{1}{M} \sum_{n=1}^{M} \gamma_n$$

- The difference ψ between the input image ϕ and the mean image was found,$\phi_i = \gamma_i - \psi$
- Next, a set of M orthonormal vectors were sought, μ_M, which best describes the distribution of the data. The k^{th} vector, μ_k, is chosen such that

$$\psi = \frac{1}{M} \sum_{n=1}^{M} (\mu_k^T \phi_n)^2$$

- λ_k is a maximum, subject to

$$\mu_l^T \mu_k = \begin{cases} 1, & \text{if } l == k. \\ 0, & \text{otherwise.} \end{cases}$$

where μ_k and λ_k are the eigenvalues and eigenvectors of the covariance matrix C.

- The covariance matrix C has been obtained,

$$\psi = \frac{1}{M} \sum_{n=1}^{M} (\phi_n \phi_n^T)^2 = AA^T$$

where $A = [\phi_1, \phi_2, \phi_3, ..., \phi_m]$.

- To find eigenvectors from the covariance matrix is a huge computational task. Since $M$is far less than N^2 by N^2, we can construct the M by M matrix,

$$L = A^T A$$

where $L_{mn} = \phi_m^T \phi_n$

- We find the M Eigenvectors, v_l of L.These vectors (v_l) determine linear combinations of the M training set face images to form the Eigenfaces u_l.

$$\mu_l = \sum_{k=1}^{M} v_{lk} \phi_k$$

where $l = 1, 2, 3, ..., M$

- After computing the Eigenvectors and Eigenvalues on the covariance matrix of the training images
  - M eigenvectors are sorted by Eigenvalues

  - Top eigenvectors represent Eigenspace

- Project each of the original images into Eigenspace to find a vector of weights representing the contribution of each Eigenface to the reconstruction of the given image.

When detecting a new face, the facial image is projected in the Eigenspace and the Euclidian distance between the new face and all the faces in the Eigenspace is measured. The face that represents the closest distance will be considered as a match for the new image. Similar process is followed for Eigenlips and Eigeneyes methods. The mathematical steps are as follows:

- Any new image is projected into Eigenspace and we find the face-key by $\omega_k = \mu_k^T$ and $\omega^T = [\omega_1, \omega_2, \omega_3, ..., \omega_M]$

  where, u_k is the k^th eigenvector and ω_k is the k^th weight in the weight vector $\omega^T = [\omega_1, \omega_2, \omega_3, ..., \omega_M]$

- The M weights represent the contribution of each respective Eigenfaces. The vector Ω, is taken as the 'face-key' for a face's image projected into Eigenspace.

- We compare any two 'face-keys' by a simple Euclidean distance measure

$$\epsilon = ||\Omega_a - \Omega_b||^2$$

- An acceptance (the two face images match) or rejection (the two images do not match) is determined by applying a threshold.

3.3.7 Energy Expenditure from Body Movement

There exists a significant correlation between accelerometer data and the work done by a person. It is found that the energy measured by ADInstrument Exercise

Phsyiology Kit is highly correlated with accelerometer energy when the phone is positioned at the waist [33].

Droid X uses the STMicroelectronics LIS331DL accelerometer. In this study, 2 Droid X 3G devices running Android OS 2.2 and one Samsung Galaxy Nexus with Android OS 3 were used as acceleration measurement platforms.

Since this is a piezo-resistive accelerometer, low pass filtering is required to acquire the true activity-component. Low-pass filtering was applied on the raw accelerometer data, as its output includes a DC gravitational contribution. In the literature, the ideal cut-off frequency or the filter ranges from 0.1 Hz to 0.5 Hz. 0.5 Hz filter was used in Matlab to exclude the gravitational contribution. After testing the varying frequency in this range, a good result was found preserving the activity contribution.

To correlate accelerometer data with energy expenditure of a person, the accelerometer's three dimensional vector needs to be summarized as one scalar value that represents physical activity intensity over small time periods [33]. This scalar value is considered accelerometer energy spent by the user. To calculate accelerometer energy, several different methods have been proposed, but the most used one is the summation of time integrals of accelerometer output over the three spatial axes [33] that has been used to calculate the accelerometer energy. The accelerometer energy is calculated according to the following formula:

$$\text{Accelerometer energy} = \int_{t_0}^{t_0+T} |a_x| + |a_y| + |a_z| dt$$

Here a_x, a_y and a_z are low-pass filtered accelerometer data corresponding to the x, y, and z axes. For calculating the values of this equation, the accelerometer input data on each of the axes were observed. Then low pass filtering was used on each axis input. Next, the absolute value of the accelerometer inputs was calculated and the integration during fifteen minute time before taking the user image was measured.

3.3.8 Fusion Using Naïve Bayes

The mean of the energy data was found for different affective states and those means were used as a separate feature for the fusion. Table 3.1 summarizes the mean of the energy for different affective states. Those means were used as the additional feature for the fusion algorithm.

Affective State	**Energy(mean)**
Anger	8.56
Disgust	22.4
Fear	41.2
Happy	15.1
Sad	4.28
Surprise	35.6

Table 3.1: Mean of energy for different affective states.

It is argued that human behavior is close to that predicted by Bayesian decision theory [55]. Different probabilistic graphical model algorithms are used in the literature like Hidden Markov Model (HMM) and Support Vector Machine(SVM).

Bayesian classifier was used for fusion of the two modalities. Since there are only two modalities, it is argued that Naïve Bayes algorithm would be a better fit, which performs better with small number of features and potentially large data for fusion . Fusing the modalities of facial expression and energy data at decision level enables to gain the knowledge about the relationship between these two modalities for a particular affective state [64].

The Bayesian fusion framework that is applied is proposed in [91]. It uses the

conditional error distributions of each classifier to approximate uncertainty about that classifiers decision. The combined decision is the weighted sum of the individual decisions. Given a problem with K classes and C different classifiers, λ_i, $i = 1, ..., C$ we like to infer the true class label ω, given the observation x. Assuming that for each classifier λ_i we have a predicted class label ω_k, where $k = 1, ..., K$ then the true class label can be derived as follows:

$$P(\omega|x) \approx P(\omega|\omega_k, \lambda_i)P(\omega_k|\lambda_i, x)P(\lambda_i|x)$$

Probabilities $P(\omega|\omega_k, \lambda_i$ and $P(\lambda_i|x)$ are used to weight the combined decision and can be approximated from the confusion matrix of classifier λ_i.

The energy expenditure data of the same persons were used from our facial expression database. When the users simulated affective state, their energy data was also collected for the last 15 minutes before taking the photograph.

3.4 Results and Evaluation

The system was evaluated in four different ways. Validating the ground truth, performance of unimodal system with only facial expression, validating energy data, and performance of the multi-modal system.

3.4.1 Validating The Ground truth

After the data collection, each day the participants were interviewed and asked about their labeling. It was found that some data were not properly labeled. Due to ambiguity of the context, some data were also discarded. For example, on one occasion PA2 said, *'I was feeling very good with my grade, but did not have much movement since I was sitting on my desk. So I labeled the emotion as positive in valence but negative in arousal and did not know which one to pick from radio button. So I selected sad.'* We only incorporated the data that the researcher and the PA we

agreed on to be of any particular affective state.

3.4.2 Unimodal System With Facial Expression

The algorithm was trained with the pictures taken by the camera of the smart phone. For each image in the training database, the classifier for facial expression was used and 89% accuracy was found. The confusion matrix for facial expression is given in Table 3.2. It was found that the pictures taken by the camera for which the environment was dark, the system gave inaccurate results and the image was not properly classified. One inaccurate results were found for each of the expression anger, sad, disgust, fear and surprise. From the result, it was concluded that the classifier works well with the training database as long as the image is taken properly with proper lighting. It did not depend on any particular expression.

a	**b**	**c**	**d**	**e**	**f**	**←Classified as**
8	0	0	0	0	0	| a=happy
0	7	0	0	1	0	| b=anger
1	0	7	0	0	0	| c=sad
1	0	0	7	0	0	| d=disgust
1	0	0	0	7	0	| e=fear
1	0	0	0	0	7	| f=surprise

Table 3.2: Confusion matrix for facial expression classifier.

3.4.3 Validating Energy Data

It was observed that the mean of the energy data for different affective states as the second feature for our Naïve Bayes fusion. Furthermore, an interesting relationship

was observed between the energy and the different emotional categories. Figure 3.3(a) shows the energy mean for different annotations by different PAs. Each point represents a particular annotation by any PA. It was difficult to visually distinguish the energy for the different categories.

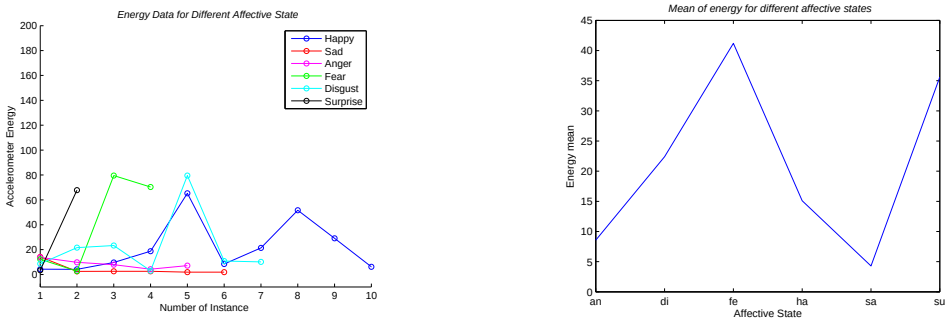


Figure 3.3: Accelerometer Energy for Different Affective State: (a) Accelerometer energy for six basic emotions. (b) Mean of energy for different affective state.

However, for the three categories, namely happy, sad and anger; an important relationship was found. The mean of energy of sad is much lower than the mean of the energy of happy and that of anger. On the other hand, the mean of energy of anger is not very high where in Russell's two dimensional space it is considered higher than the happy state. It was concluded from the data that happiness usually has a high energy expenditure relative to sadness, which is in line with Russell's theory.

This relationship is best shown in Figure 3.3(b), where only the mean for different affective states are plotted. The horizontal axis represents different emotional states and the vertical axis represents the energy mean for the corresponding emotion. It was found that sadness has much lower mean of energy than that of happiness. Also, fear has high value in arousal space and the mean was found to be much higher than happy and sad. This is also in line with the Russell's circumplex model where fear is phrased as afraid [Figure 3.1].

a	**b**	**c**	**d**	**e**	**f**	←**Classified as**
8	0	0	0	0	0	| a=happy
0	8	0	0	0	0	| b=anger
1	0	7	0	0	0	| c=sad
0	0	0	7	0	1	| d=disgust
0	0	0	0	8	0	| e=fear
0	0	0	1	0	7	| f=surprise

Table 3.3: Confusion matrix using Naïve Bayes classifier.

3.4.4 Performance of Multi-modal System

The last part of the discussion addresses the performance of the multi-modal system. The system performs better with the Naïve Bayes fusion. The system performance for correctly classifying the instances for the training database increased from 89% to 93%. Out of 48 total instances, 45 were classified correctly.

A close analysis from the confusion matrix of the multimodal system from Table 3.3 gives the rationale for the improvement.

First, the image previously misclassified as fear instead of anger is classified correctly now. The reason is that the mean of energy for anger (8.56) is much lower than that of fear (41.2). As a result, even if the image was not clear enough, it is correctly classified. The same reasoning is also true for the data that was previously mis-classified as anger instead of fear.

This data is now also classified correctly. Another interesting observation is the confusion matrix entry for disgust. One disgust entry was mis-classified in the unimodal system as anger.

In the multi-modal system it still is mis-classified, but to a different class, fear. It is observed that the mean of energy for disgust is equidistant from both anger and fear. As a result, the system could not find a close match for this annotation.

3.5 Findings and Discussion

3.5.1 Inherent Theory of Emotion is Not Established

The theory of emotion is not established yet. Psychologists have different approaches to identify different emotions. Research in the field of affective computing is about finding the features that are most likely related to emotion-oriented computing. Understanding those ideas and adapting those to any computational methods is still in progress. Furthermore, expression of emotion greatly varies from person to person, man and women, and also among different age groups and races.

Paul Ekman has identified six basic emotions for psychologists to identify from video sequence using Facial Action Coding System (FACS). Those are happiness, sadness, disgust, anger, fear and surprise. There are other emotions important for automatic detection of emotions like boredom, frustration, excitement and many more. Even Ekman expanded his list of basic emotions to include other emotions like amusement, contempt, embarrassment, excitement, guilt, satisfaction etc.

There are also different approaches in computing for different theories in psychology. For Ekman's FACS to be implemented; feature extraction is needed from facial image and then it needs to be classified. However, there are different emotions with overlapping Coding Schemes which makes the implementation complicated.

For the holistic approach different machine learning algorithms are used. We have used such an approach.

3.5.2 Multimodal System Needs More Modalities

We have argued that multimodal emotion recognition will contribute to the more accurate affective classification. For that we might have to put different weights

for different modalities. Also, in person to person communication, we may or may not look for the same features in multiple channels like facial expression, speech and body movements. More importance might be needed for finding same emotional cues in multiple modalities. Again, this varies a lot among person to person. People tend to understand about others emotion from facial expression, tone, body movement, gestures and most importantly context. Depending upon context, the interpretation of a message could be quite different from another. A combination of low level features, high level reasoning, and natural language processing is likely to provide best multimodal affect recognition. But very few systems have been developed in a natural environment considering multiple modalities. Even if they were developed, their performance is measured in a laboratory environment, which might be quite different than in a natural environment.

3.5.3 Privacy

We have argued that affect detection is important but that also comes with increasing concerns about privacy awareness of the people. However, this argument can be contrasted with the fact that in our system, detected affective state is shared only by the permission of the user. Nevertheless, there remains significant scope for research regarding privacy issues and different levels of anonymization techniques to be dealt with.

3.6 Conclusion

We provided much attention to validate the ‘ground truth’ data, we found that some emotional states are ambiguous and even human can not identify the emotional states properly. This is because human might have mixed emotions at a particular time. There are no borders with different affective states. However, still we emphasized on the labeling of the emotion by the PAs. Depending upon the interview with the researcher, we incorporated that data for our training database or not. The success of

the system largely depends on the emotional self-awareness of the PAs. We think that instead of a particular classification of a particular affective state, one particular instance should be labeled with the different probabilities of falling into different categories. Depending upon those probabilities, machine interpretation of affective states can be applied to human computer interaction. Also, affect sensitive applications should be developed targeting the application scenario. For example, the application for advertisement in smart phones may not be feasible for detecting boredom in a learning environment. We also find that arousal can be captured easier than valence. One such application might be capturing arousal from pupil size which is also a good approximation of the arousal space. However, for mobile devices it might not be appropriate because of the change of lighting and all other conditions. We believe that by real time sensing of affective states using smart phones, we can machine interpret human affective states and machines can understand part of larger human intelligence. With the continuous advancement of sensor technologies in smart phones, we can predict human affective states more accurately and the application of such affect detection technique might be huge.

3.7 Related Publications

- **Mohammad Adibuzzaman**, Niharika Jain, Nicholas Steinhafel, Munir Haque, Ferdaus Kawsar, Richard Love, Sheikh Iqbal Ahamed: ***Towards In Situ Affect Detection in Mobile Devices: A multimodal approach,*** in Research in Adaptive and Convergent Systems(RACS, 2013)

- **Mohammad Adibuzzaman**, Niharika Jain, Nicholas Steinhafel, Munirul Haque, Ferdaus Ahmed, Sheikh Ahamed, Richard Love: ***In-situ Affect Detection in Mobile Devices: A Multimodal Approach for Advertisement Using Social Network*** in ACM SIGAPP Applied Computing Review,13(4): 67-77 (2013)

- Md. Munirul Haque, **Md Adibuzzaman** et al.:***IRENE: Context Aware Mood Sharing for Social Network,*** in KASTLES 2011 Workshop.

Chapter 4

Pain Level: Smart Phone Based Pain Level Detection

4.1 Introduction

Timely and accurate information about patients symptoms is important for clinical decision making such as adjustment of medication. Because of the limitations of self-reported symptoms such as pain, whether facial images can be used for detecting pain level accurately using existing algorithms and infrastructure for cancer patients was investigated . For low cost and better pain management solution, a smart phone based system for pain expression recognition from facial images is presented. To the best of our knowledge, this is the first study for mobile based chronic pain intensity detection. The proposed algorithms classify faces, represented as a weighted combination of Eigenfaces, using an angular distance, and support vector machines (SVMs). A pain score was assigned to each image by the subject. The study was done in two phases. In the first phase, data were collected as a part of a six month long longitudinal study in Bangladesh. In the second phase, pain images were collected for a cross-sectional study in three different countries: Bangladesh, Nepal and the United States. The study shows that a personalized model for pain assessment performs better for automatic pain assessment and the training set should contain varying levels of pain representing the application scenario.

In excess of 8 million individuals globally die each year from cancer and three-quarters of these are reported to suffer from pain [99]. A primary barrier to provision of adequate symptom treatment is failure to appreciate the intensity of the symptoms most commonly pain–patients are experiencing [88]. One difficulty for health care providers in helping patients with chronic conditions like cancers is having accurate, complete, and timely information about symptoms, daily information if possible. In particular, failure to use (repeatedly) validated symptom assessment tools

prevents communication between patients and health-caregivers to bring attention to symptoms issues [9]. The usual way to obtain such information is to ask office-visiting patients standard questions about their symptoms and their intensities. For patients with cancer the most widely used questionnaires for this task are the Edmonton Symptom Assessment Survey (ESAS) or the Brief Pain Inventory [20][24]. Common practice is to have patients provide answers on paper to these instruments when they are seen in doctors offices. This practice of course means that the data obtained only cover the particular situation the patient is in at that time. For example the patient may have taken extra pain medicines because of the appointment trip and wait in the doctors office. A more abbreviated symptom assessment strategy in doctor-patient encounters is simply to have patients verbally report their current level of pain on a visual analogue 1-10 scale; sometimes a picture of this scale with figure faces showing different levels of distress is used. However, this is a one-time and one-item assessment strategy.

In an ideal situation, to monitor patients more completely and know every day how patients feel and then of course to make adjustments in treatments, such as types, amounts and timing of pain medicines, it would be good to have data from such questionnaires every day. In studies where patients have home computers, such daily assessments reported by email and the resultant treatment adjustments, are associated with increased quality of life and survival in terminally-ill patients with lung cancer. In settings where hospice programs are available, patient and family satisfaction is clearly related to intensity of monitoring and consequent associated prompt adjustments of symptomatic managements. This intensity of hospice monitoring is almost always greater than patients have had in their regular care, which has been usually and mostly based on patient office visits face-to-face. Management through phone contact, or email contact is usually limited, mostly because doctors are uncomfortable with their command of the full picture of the problems they are managing, but also because the

practice of medicine has historically been based on face-to-face encounters. All of these issues are magnified in low- and middle income countries where limited access to care, sub-optimal quality of care and usually no hospice care at all, are the norms.

One more practical way to make obtaining such more detailed symptom information possible and usable by physicians, is to put the questionnaires on a cell phone software platform, which the patient or his/her attendant could then complete at home and send by phone each day to a doctors records/office. Such a system was developed [43] and we are now trying to scale this up into a tele-home hospice system in rural Bangladesh. Our experience has highlighted two broad issues; first particularly with respect to chronic pain which characterizes the situation for patients with incurable cancers, there is an apparent dis-connect between what patients report about their pain levels using the standard instruments and their affect. Patients often report high pain levels while smiling. Second, many patients cannot use these instruments because of limited cognitive abilities or other medical conditionsfor example patients who are seriously ill in medical intensive care units. These issues have led us to investigate whether we could reproducibly and accurately record and quantitate patients pain levels using cell phone camera images of facial expression. In this communication we address three broad questions in our investigation: First, can in fact facial images be used to reproducibly assess pain intensity among cancer patients? Second, what algorithm can define pain intensity most accurately? Third, what system design issues arise in this work?

Facial expression for quantitative pain assessment has its roots in psychology. There have been historical concerns about the objectivity of self-report assessments and their susceptibility to behavioral bias [77]. As a result, there is ongoing work to identify universal cues for pain expression. Prkachin et al. showed that indices of facial expression change due to variations of pain [78]. Ekman and Friesen's [30] Facial Action Coding System (FACS) has been used to identify universal Action Units

corresponding to pain expression. It has been shown that for cold, pressure, ischemia and electric shock there are significant changes in four types of actions - brow lowering, eye orbit tightening, upper-lip raising/nose wrinkling and eye closure [76]. Prkachin and Solomon defined a Prkachin and Solomon Pain Intensity (PSPI) measure as the sum of the intensities of these four actions [60]. Therefore, it appears that there are multiple facial expression components (i.e. action units) that together comprise the facial expression depicting the intensity of pain experienced by individuals.

We conclude that a few specific action units correspond to pain intensities for most pain expressions. For acute and chronic pain, the change in different action units might be of different magnitudes. A large amount of data for a particular target application (e.g. pain monitoring for chronic pain) improves the accuracy of pain intensity predictions using facial expressions [60]. While it is comparatively easy to identify changes in action units due to acute pain, it is more difficult to find the exact change in action units due to chronic pain. In this context, instead of using the FACS, we use principal component analysis to extract information that would give us reasonable variance across a given data set for pain expression. The fact that the principal component analysis gave that the four core action units comprises 0.30 or greater fraction of pain expression across all pain tests [80] supports this claim. Therefore, we chose to use a principal component based method for detection of pain expression. Eigenface method is such a method. Each of the Eigenfaces corresponds to different levels of variance in the training data set.

To the best of our knowledge, our study is the first that does both longitudinal and cross sectional study to measure pain intensity from facial images for chronic pain. There has been works to measure pain from facial expressions using principal component analysis (PCA)[67] but the accuracy was low. Previous works do not use images of a single subject over a long period using a longitudinal data set for chronic pain. A longitudinal dataset reduce the need for requiring a large sample by removing

individual differences in pain expression and behavioral bias of self-report pain intensity. In addition, a longitudinal study would allow us to know if a person's use of facial expressions for representing pain changes over time or they consistently use the same expressions. This work explores machine learning techniques for automatic pain detection from facial images. We show that an algorithm that uses Eigenface method with proper distance measure works for pain intensities.

4.1.1 Contribution

The contribution of this work is three fold. First, in this work we have designed and deployed a mobile based system for remote monitoring of pain intensity from facial images using a smart phone camera. In contrast to other systems for emotion or pain detection, we have deployed the healthcare tool and it has been tested in a clinical setting (results section). Second, we have identified the issues regarding training and testing for better algorithm development for pain intensity detection (discussion section). Third, we have identified the design challenges to overcome the barriers to deploy the system for remote monitoring of pain intensity from facial expression (design issues section). Each pain image is labeled with a value between 0 and 10 (0 being the lowest and 10 being the highest pain intensity) and the Eigenface method that we develop has a mean absolute error of 2.483 for a single training database of all the subjects. The system can be used in different settings including:

- Remote monitoring of pain
- ICU Patients
- Neonates
- Verbally impaired patients

4.2 Our Approach

4.2.1 Related Work

Self-report of pain is considered an objective measure with a behavioral bias [77]. There has been much research on obtaining a more objective measure for pain detection using tissue pathology, imaging procedures, testing of muscle strength, etc. [97]. These approaches are not suitable for our case as these are highly invasive and require specialization and also not very accurate.

Facial expressions have also been used to detect pain [60]. Research in pain detection from facial expressions was initiated by identifying the Facial Action Coding Systems (FACS) for pain [30]. Prkachin et al. validated the use of several facial action units (AUs) namely brow lowering, eye orbit tightening, levator contraction and eye closing to constitute pain expressions in the face [80]. However, the method has been recently developed and only been tested in a small sample of patients. This method was developed using a sample that combined individuals with both acute and chronic pain but it is well known that the clinical presentation of chronic pain can be very different from acute pain [41].

Pain detection from facial images using FACS has been explored [7]. But most systems that use FACS identify presence versus absence of pain, which does not provide a quantitative estimate of the pain intensity. Recently there has been a shift in research of affect detection from a lab environment to a natural environment i.e., in situ measurement of emotion. Ayzenburg et al. [10] used electro dermal activity for stress detection of a mobile user in their natural environment. Microsofts SenseCam has also been used for logging emotional data in a natural setting. Isbister and her colleagues designed Yamove! to introduce social interaction in dance games [47].

Many research studies describe facial expressions that individuals use when in pain at different age levels (neonatal, adult) [27][28][29][37][38], groups of patients

including pregnant women, patients suffering from shoulder pain [79], knee pain, and chronic pain etc. However, limited work has been conducted on the topic of automatic detection of pain from facial expressions. Several studies have been performed by examining which facial action units are used to express pain [59][58].

A number of methods have been used to identify pain from facial expressions. Different classifiers such as support vector machine (SVM), AdaBooster, Gabor filter, and hidden Markov model (HMM) have been used alone or in combination with others to achieve greater accuracy. Researchers have used active appearance models (AAM) to identify specific facial features associated with pain [8][7]. The Eigenface-based method was deployed in an attempt to find a computationally inexpensive solution [66]. Later the authors included Eigeneyes and Eigenlips to increase classification accuracy [68]. Several authors have relied on artificial neural network-based back propagation algorithms to distinguish between pain versus no pain from facial features [83][67]. A Bayesian extension of SVM named relevance vector machine (RVM) has been adopted [71] to increase classification accuracy. Niese et al. and Becouze et al. attempted to measure pain experienced by ICU/post-operative patients [12][71]. Becouze et al. built an algorithm based on the hypothesis that pain results in forming extra wrinkles [12]. Niese et al. used a photogrammetric technique for finding features and later a SVM filter with a radial basis function (RBF) Gaussian kernel was for detecting pain [71]. Brahnam et al. worked to find pain in neonates [19][18].

Monwar et al. presented a method for recognizing pain expression from video [66]. The extracted face image was used to build a feature space defined by Eigenfaces. The Eigenfaces are images corresponding to different eigenvectors of the training set images. Finally, each new image was tested by projecting the image onto the feature space and finding the corresponding label of the training set. The closest labeling was found using the Euclidean distance measurement of the weight vector.

Ashraf et al. also presented an approach for detecting pain from video

sequences of facial expressions [8]. The image database was obtained from the UNBC-McMaster shoulder pain expression database with images of adult patients with rotator cuff injuries. The authors used a holistic approach to create active appearance models (AAM) for a face and support vector machines (SVM) were used as the classifier.

Boehner et al. [15] showed that emotion is not an objective measure and should be considered as a subjective experience. Pain is obviously a subjective form of emotion and should also be measured out of the lab. The Eigenface method was applied in face recognition in [98]. It has also been applied in emotion recognition [105] and pain and no pain condition detection in face images [66].

Almost all of these approaches are impacted by one or more of the following deficits: 1) reliability on a clear frontal image, 2) out-of-plane head rotation, 3) dilemma in correct feature selection, 4) failure to use temporal and dynamic information, 5) considerable amount of manual interaction, 6) inability to handle noise, illumination, glass, facial hair, skin color issues, 7) high computational cost, 8) lack of mobility and 9) failure to classify intensity of pain level.

4.2.2 Design Issues

In the beginning of our work, we spent several weeks in clinics and hospitals and in home visits with cancer patients in Bangladesh learning about patient needs and health care professional challenges. Our experiences reinforced our sense that the usual current system of office visits for patients with advanced cancers and pain serves patients poorly. Physicians expressed significant interest in having real time usual day and activity symptom data on their patients. Subsequently we spent similar durations of time in the field working on deployment of our system.

Availability of Mobile Network

Among poor rural Bangladeshi patients we were encouraged to find that among 45 patients surveyed, 43 had access to a cell phone. Additionally we found that these patients were served by good data networks. Most image processing techniques require very high computing power and it is difficult to use a smart phone for this purpose. The availability of good mobile data networks made it possible for us to use the cloud for photo images using advanced software such as Matlab.

Smile for the camera

Common practice in affective computing is that when a person has a photo image of their face taken for assessment of affective state, he or she is asked to pose for the camera or smile for the camera [15]. To address the biases this usual practice likely creates, we elected to take two facial expression pictures in each studied patient. The first (candid) image was taken without giving any instructions. The second image was taken after providing specific directions. We instructed the patient to make a facial expression that reflected their current pain level. We defined that image as the acted pain and the image without instruction as the real pain for the purpose of this paper. In our training data set to be described below (dataset subsection), the images were randomly selected from acted pain and real pain images.

Personalized Model

In the first phase of our work, our goal was to identify pain levels from facial photographs for individuals followed longitudinally. There is a significant amount of variance in pain levels with facial expressions among individuals. As a result, we hypothesized that a training database with multiple images from a single person would eliminate this variance. Furthermore, individualized training databases for pain intensity would accommodate for different total numbers of images from individuals.

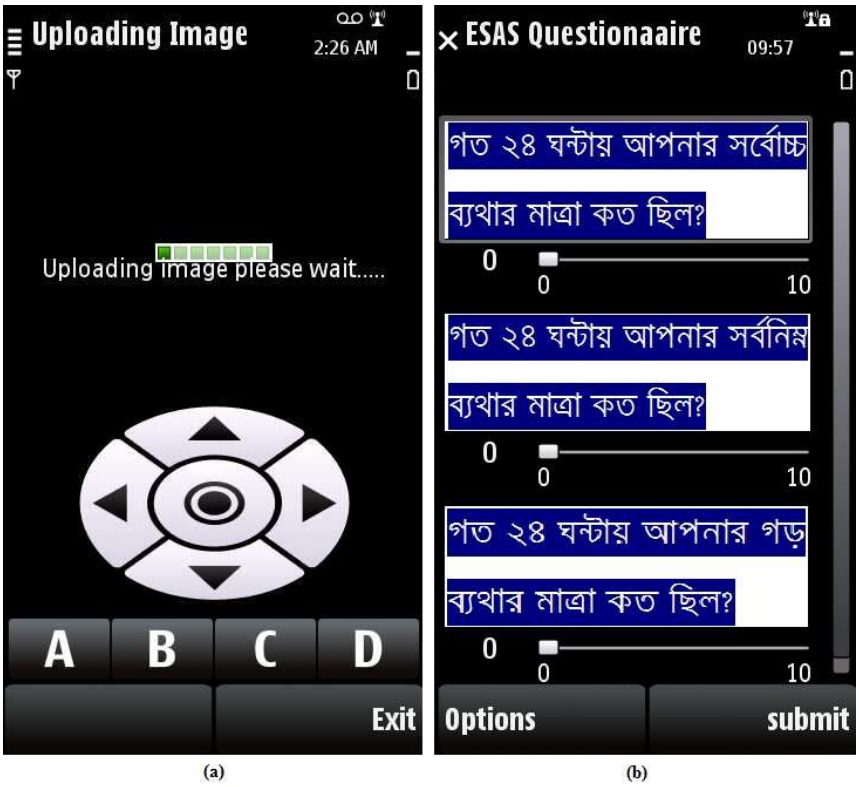


Figure 4.1: Screen shots of mobile application for labeling and uploading pain image. The first screen (a) shows status of image upload and the second screen (b) shows labeling of pain intensity using a sliding bar in the local language Bangla.

4.2.3 Data

Our protocol for the longitudinal pilot study in a small number of patients was approved by the Institutional Review Board (IRB) at Marquette University and by The Bangladesh Medical Research Council in Bangladesh. All patients provided written informed consent. Each patient was given a Nokia X6 phone with internet provided by Grameen Phone, the largest mobile service provider in Bangladesh. The patients (all women) were aged between 35 and 48 years.

Fourteen (14) subjects were recruited for the study. Each subject and the attendant of the subject were trained how to take the pictures using the camera at the health center. The key aspects of the training and image creation are:

- A doctor would take a picture of the subject.

- The subjects were told about the Visual Analog Scale (VAS), which describes the pain intensity with 0 being the no pain and 10 being maximum pain possible.

Figure 4.2 was used to explain the scale of pain intensity which is also written in local language Bangla. While taking the picture, the subjects were asked

- To take a picture of only the face.
- To use a light background.
- Not to use sunglasses, if possible.

From among the 14 patients, 6 lived longer than 3 months and regularly provided a total of 454 usable images

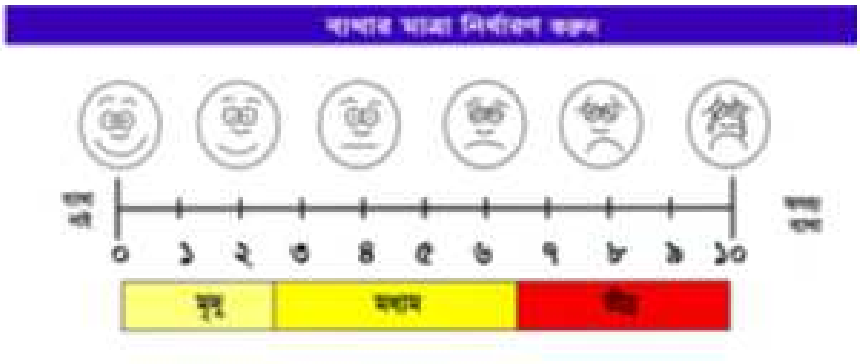


Figure 4.2: Visual Analog Scale (VAS) in the local language Bangla used for training. For each of the pain intensity levels indicated on the line, the image of the face represents possible expression in the face. 0 means no pain and 10 (far right) indicates the highest possible level of pain.

In the second phase of our work we conducted a cross sectional study. The protocol for this study was approved at Marquette University and by the responsible ethical review boards in Bangladesh, Nepal and Rapid City South Dakota in the United States. In this study we recruited patients presenting for a clinic visit with advanced cancer and at that single visit obtained two facial images as noted aboveone candid and one after specific instructions. Table 4.1 shows that we obtained usable photographs

Longitudinal Study			
Subject	Training Set	Test Set	Total
A	6	8	14
B	36	80	116
C	36	124	160
D	6	6	12
E	36	78	114
F	6	32	38
Cross-sectional Study			
Location		Training Set	Test Set
Bangladesh		454	131
Nepal		454	311
United States		454	513

Table 4.1: Image data set size for longitudinal and cross sectional study. The entire data set for longitudinal study was used as the training data set for the cross sectional study.

for 131 Bangladeshi, 311 Nepali, and 71 American Indian patients. 36 randomly selected images were used as the training set for each subject during the longitudinal study. For the cross sectional study, the entire dataset of the longitudinal study (454 images) was used as the training set.

4.2.4 Eigenfaces

This method is based on Principal Component Analysis (PCA) [95][96]. The analysis technique includes only the characteristic features of the face corresponding to a specific pain level and leaves other features. The resultant images are used as samples for training with M Eigenfaces for the highest Eigenvalues. The Eigenface method was described in chapter 3.

This technique can be explained using the Figure 4.3. The projected image is a linear combination of eigenfaces. We find the image that is the closest match of the linear combination of the eigenfaces.

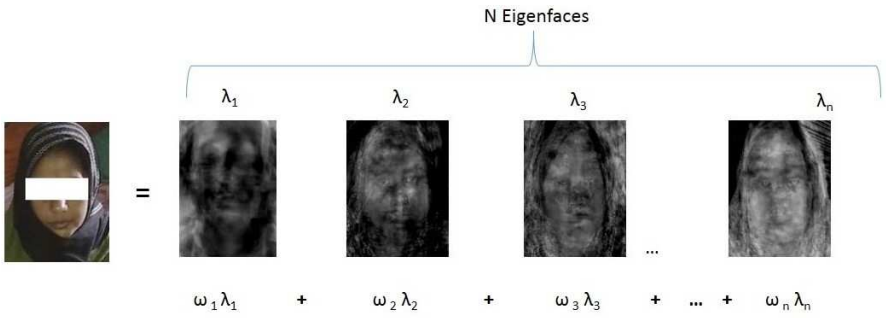


Figure 4.3: Each image is represented as a linear combination of the Eigenfaces

4.2.5 Closest weight vector of the image

For the classification of the weight vectors, we applied three approaches. These approaches used Euclidean distance, angular distance and support vector machine respectively. Different distant measures were needed for the high dimensions of the weight vectors. The dimension of the weight vector is equal to the number of images minus one in the training set. Consequently, for a training set of 36 images, we have 35 Eigenfaces and the weight vector was of 35 dimensions. The angular distance defined in [103] is, $d(A, B) = \frac{(A.B)}{(|A||B|)}$ Angular distance works better (mean absolute error decreases) than Euclidean distance in high dimensional space. We also used support vector machine to improve the sensitivity and specificity for each pain class. Support vector machine is a classification problem where the series of input variables X_1, X_2, X_3X_n and their corresponding class labels C_1, C_2, C_3C_k are given. The classification problem is given a new input variable X, which would be the class label. This is a binary classification problem which can be extended to multi-class classifier. To classify a new input vector, the classifier function is defined as $y(x) = W^T\psi(x) + b$ Where ψ is a continuous feature space transformation, W is the weight vector and b is the bias parameter.

4.3 Results and Evaluation

The results of our system were evaluated in terms of two performance measure: the mean absolute error and, sensitivity and specificity analysis for the three pain classes, low (L), medium (M) and high (H). Pain level between 1 and 4 was termed as low, between 5 and 7 was considered as medium and between 8 and 10 was defined as high. This classification into three categories is similar to the Brief Pain Inventory which has been proposed and validated across different cultures [24]. As a consequence of insufficient data for subjects A, D, and F (Table 4.1), the results are shown only for subjects B, C and E, for which there were 36 images in the training set.

4.3.1 First phase–longitudinal study

Mean Absolute Error

We have six different training sets for the six subjects from the first phased longitudinal study. These training sets have a randomized combination of 36 images when available (subjects B, C, E) of acted and real pain. The training sets in this setup are referred as personalized training database in this paper. There are two reasons for this. As indicated earlier such a training set would eliminate the individual differences in pain expression. Second, our 'gold standard' or ground truth, the pain level provided by the subjects, is objective but with a behavioral bias. A personalized training database would eliminate that behavioral bias. Each subjects images were tested against the training set of the corresponding subject. With the personalized training database, we tested the classification algorithm with three distant measures: Euclidean distance, angular distance and support vector machine. Euclidean distance gave poor results and is not reported here. The mean absolute error for angular distance and support vector machine is shown in Table 4.2. Subjects A, D and F had only six images in the training set. As a result, the method did not work well and the results for subjects B, C and E are reported here. A 10 fold cross validation was performed.

	Subject B		Subject C		Subject E	
Cross Val	Angular	SVM	Angular	SVM	Angular	SVM
1	0.95	1.07	0.71	0.88	1.06	0.64
2	1.02	1.142	0.71	0.77	1.01	0.67
3	0.79	0.81	0.75	0.80	1.04	0.68
4	1	1.01	0.8	0.78	0.98	0.66
5	1.12	0.97	0.83	0.83	0.98	0.72
6	1.07	0.86	0.707	0.94	1.22	0.66
7	0.88	0.94	0.82	0.87	1.09	0.62
8	0.83	0.91	0.73	0.92	1.12	0.75
9	0.92	0.73	0.78	0.82	1.04	0.53
10	1.04	1.05	0.79	0.78	0.96	0.63
Mean $\pm$SD	0.96 $\pm$ 0.10	0.94 $\pm$ 0.12	0.76 $\pm$ 0.04	0.84 $\pm$ 0.06	1.05 $\pm$ 0.08	0.66 $\pm$ 0.05

Table 4.2: Mean absolute error for a 10 fold cross validation for the longitudinal study.

Sensitivity Analysis

For reproducible use in clinical settings it would be optimal to reduce the mean absolute error for pain level assessment. It is also important that the input and output pain distributions are similar or it may be possible that the system is always giving the same pain level as output but the mean absolute error is low. From a machine learning perspective, the system would perform well when the input pain level distribution is similar to that of the training data set. For a robust clinical decision support system we want the system performing well independent of the input pain level distribution. The sensitivity and specificity of each class (low, medium and high) of a 10 fold cross validation for subjects B, C, and E are shown in Figure 4.4.

	Angular						SVM					
Subject	Sensitivity			Specificity			Sensitivity			Specificity		
B	L(0-4)	M(5-7)	H(8-10)	L(0-4)	M(5-7)	H(8-10)	L(0-4)	M(5-7)	H(8-10)	L(0-4)	M(5-7)	H(8-10)
	0.25	0.9	NaN	0.9	0.25	1	0.06	0.92	NaN	0.92	0.06	1
	0	1	NaN	1	0	1	0.16	0.88	NaN	0.88	0.16	1
	0	1	NaN	1	0	1	0.42	0.84	NaN	0.84	0.42	1
	0	1	NaN	1	0	1	0.13	0.84	NaN	0.84	0.13	1
	0.86	0.54	NaN	0.54	0.86	1	0.14	0.91	NaN	0.91	0.14	1
	0.43	0.74	NaN	0.74	0.43	1	0.14	0.94	NaN	0.94	0.14	1
	0	1	NaN	1	0	1	0.07	0.95	NaN	0.95	0.07	1
	0	1	NaN	1	0	1	0.33	0.76	NaN	0.76	0.33	1
	0.24	0.89	NaN	0.89	0.24	1	0.06	0.94	NaN	0.94	0.06	1
	0	1	NaN	1	0	1	0.27	0.92	NaN	0.92	0.27	1
Mean	0.18	0.91	NaN	0.91	0.18	1	0.18	0.89	NaN	0.89	0.18	1
C	1	0	NaN	0	1	1	0.95	0	NaN	0	0.95	1
	1	0	NaN	0	1	1	0.97	0	NaN	0	0.97	1
	1	0	NaN	0	1	1	1	0.1	NaN	0.1	1	1
	1	0	NaN	0	1	1	1	0	NaN	0	1	1
	1	0	NaN	0	1	1	1	0	NaN	0	1	1
	1	0	NaN	0	1	1	0.89	0	NaN	0	0.89	1
	1	0	NaN	0	1	1	0.92	0.13	NaN	0.13	0.92	1
	1	0	NaN	0	1	1	0.96	0	NaN	0	0.96	1
	1	0	NaN	0	1	1	1	0	NaN	0	1	1
	1	0	NaN	0	1	1	0.97	0.14	NaN	0.14	0.97	1
Mean	1	0	NaN	0	1	1	0.97	0.04	NaN	0.04	0.97	1
E	0.07	0.9	NaN	0.9	1	1	0.21	0.95	NaN	0.95	0.21	1
	0.07	1	NaN	1	0.07	1	0.14	0.97	NaN	0.97	0.14	1
	0	1	NaN	1	0	1	0.07	0.98	NaN	0.98	0.07	1
	0.07	1	NaN	1	0.07	1	0.21	0.97	NaN	0.97	0.21	1
	0	1	NaN	1	0	1	0.23	0.95	NaN	0.95	0.23	1
	0.23	0.61	NaN	0.61	0.23	1	0.23	0.97	NaN	0.97	0.23	1
	0.13	0.88	NaN	0.88	0.13	1	0.25	0.98	NaN	0.98	0.25	1
	0	0.87	NaN	0.87	0	1	0.31	0.97	NaN	0.97	0.31	1
	0.45	0.65	NaN	0.65	0.45	1	0.27	0.97	NaN	0.97	0.27	1
	0.1	0.89	NaN	0.89	0.1	1	0.5	0.95	NaN	0.95	0.5	1
Mean	0.11	0.88	NaN	0.88	0.21	1	0.24	0.97	NaN	0.97	0.24	1
Mean ± SD	0.43± 0.45	0.60± 0.44	NaN	0.60± 0.44	0.46± 0.45	1 ± 0	0.46± 0.37	0.60± 0.43	NaN	0.63± 0.43	0.46± 0.37	1 ± 0

Figure 4.4: Sensitivity and specificity of a 10 fold cross validation for the longitudinal study

4.3.2 Second phase– cross-sectional study

Our findings from the first phase analyses could benefit from validation across a large numbers and different populations. Because of the individual differences in pain expression and behavioral bias from self-reported pain level data, a decrease in the system performance with non- longitudinal and different population data was expected. For the cross sectional study we had one image for each subject with a total

Angular						SVM					
Sensitivity			Specificity			Sensitivity			Specificity		
L	M	H	L	M	H	L	M	H	L	M	H
0.55	0.39	0.02	0.40	0.58	0.99	0	1	0	1	0	1

Table 4.3: Sensitivity and specificity for the cross-sectional study when the entire data set from the longitudinal study was used as the training data set.

of 513 subjects. We experimented with different data set for the training. We found that when we used the entire data set for the longitudinal study (454 images for six subjects) as the training database, we had a mean absolute error of 2.91. The sensitivity and specificity analysis is given in Table 4.3.

4.4 Discussion and Findings

4.4.1 Personalized model works better

The classification accuracy using the method works much better for the longitudinal study when we use the images of one person over a long time. Table 4.2 shows that we had a mean absolute error less than 1 for the longitudinal study. This proves the subjectivity of pain expression and reflects the behavioral bias of the individual. We also found for the Eigenface method, angular distance and SVM gave similar result when used with the longitudinal dataset, but angular distance was better for the cross-sectional study (Table 4.3).

4.4.2 Distribution of pain level in the training set

The sensitivity and specificity varied much across different subjects and different training database. The primary reason for that is the lack of images of the representing class (low, medium and high) in the training data set. For example, the sensitivity was very high for the class low for subject C whereas for subject E, the sensitivity was high for the medium pain level across each cross validation (Figure 4.4). Analysis of the percentage of images with low and medium pain level in the training database explains this result. We had a high percentage of images with low

pain level for subject C in the training set, resulting in better accuracy for low pain level for subject C (Figure 4.5).

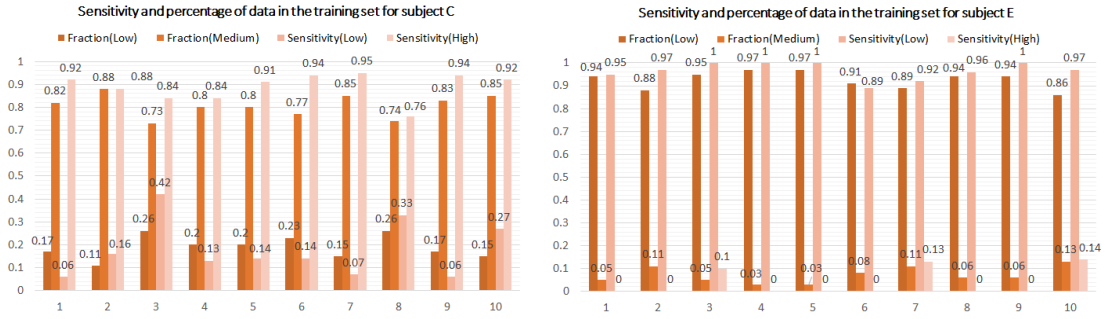


Figure 4.5: Fraction of the number of images for the two different classes (low and medium) and the sensitivity for each class for the 10 fold cross validation during the longitudinal study.

4.4.3 Application scenario

It has been shown that pain measurement can be used in clinical settings for the improvement of the quality of life for cancer patients with pain using pain assessment tools such as Brief Pain Inventory (BPI) [5]. Consequently, automatic pain detection into three categories: low, medium and high has application in clinical settings. The primary goal is accurate and timely intervention for cancer patients with pain. One of the barriers to that is inadequate measurement of pain levels and such systems are of importance to address this problem.

4.5 Conclusion

Automatic emotion detection from facial images is a challenging research problem and a significant amount of work has been done in this area during the last twenty years [96][104]. The success of these application specific systems depends on narrowing down the context of the application and collecting enough data from specific settings [60]. In this paper we showed that a smart phone based tool can be used for

remote monitoring of pain intensity for long term pain management with appropriate training dataset. Most works for pain detection involve the detection of pain or no pain. But in a clinical setting pain intensity is very important. The use of a mobile phone for pain intensity detection might reduce the healthcare costs and allow assessments in otherwise un-evaluable patients. Further work is needed to address the issues of appropriate training set for target application and selection of the right algorithms. The usability of such systems with patients with chronic pain and the effect on the system performance due to candid image and acted image also needs to be investigated.

4.6 Related Publications

- Richard R. Love, Tahmina Ferdousy, Bishnu D. Paudel, Shamsun Nahar, Rumana Dowla, **Mohammad Adibuzzaman**, Golam Mushih Tanimul Ahsan, Miftah Uddin, Reza Salim, Sheikh Iqbal Ahamed: ***Symptom levels in care-seeking Bangladeshi and Nepalese adults with advanced cancer*** submitted to the Journal of Pain and Symptom Management.

- **Mohammad Adibuzzaman**, Colin Ostberg, Sheikh Iqbal Ahamed, Richard Povinelli, Bhagwant Sindhu, Richard Love, Ferdaus Kawsar, Tanimul Ahsan: ***Assessment of Pain Using a Smart Phone*** in Proceedings of The 39th Annual International Computers, Software & Applications Conference (COMPSAC 2015).

Chapter 5

Arterial Blood Pressure: Novel index for Identifying Early Markers of Hemorrhage

5.1 Introduction

Identifying the need for interventions during hemorrhage is complicated due to physiological compensation mechanisms that can stabilize vital signs until a significant amount of blood loss. Because the physiological systems providing compensation affect the arterial blood pressure waveform through changes in dynamics and waveform morphology, we propose that Markov chain analysis of the arterial blood pressure waveform can be used to monitor physiological systems changes during hemorrhage. Continuous arterial blood pressure recordings were made on anesthetized swine (N=7) during a 5 min baseline period and during a slow hemorrhage (10 ml/kg over 30 min). Markov chain analysis was applied to 20 sec arterial blood pressure waveform segments with a sliding window. 20 ranges of arterial blood pressure were defined as states and empirical transition probability matrices were determined for each 20 sec segment. The mixing rate (2nd largest eigenvalue of the transition probability matrix) was determined for all segments. A change in the mixing rate from baseline estimates was identified during hemorrhage for each animal (median time of 13 min, 10% estimated blood volume, with minimum and maximum times of 2 and 33 min, respectively). The mixing rate was found to have an inverse correlation with shock index for all 7 animals (median correlation coefficient of -0.95 with minimum and maximum of -0.98 and -0.58, respectively). The Markov chain mixing rate of arterial blood pressure recordings is a novel potential biomarker for monitoring and understanding physiological systems during hemorrhage.

5.2 Related Work

Hemorrhage is a medical emergency frequently encountered by clinicians in situations as diverse as emergency and operating rooms, intensive care units or mass casualty incidents. A significant amount of blood loss due to hemorrhage can cause hemodynamic instability, inadequate issue perfusion, hemorrhagic shock, and, if left untreated, eventual death [39]. Hemorrhage is the cause of 40% of deaths after a traumatic injury in the United States [54]. One of the limitations to treating hemorrhage is that vital signs can appear normal until a significant blood loss has occurred. This delay in vital sign changes is due to the action of the sympathetic and parasympathetic control of blood pressure, which can effectively compensate until blood loss is significant. There is therefore much interest and value in identifying early and sensitive biomarkers of hemorrhage.

Heart rate variability is one method suggested in the literature to identify hemodynamic instability due to hemorrhage [94]. Although it has been shown that aggregate group mean values of heart rate variability are correlated with stroke volume, heart rate variability is less reliable when tracking individual reductions in central volume during progressive lower body negative pressure or simulated hemorrhage [87]. It has been suggested that reductions in vagal activity assessed with heart rate variability or baroreflex sequences may represent identifiable early markers of hemorrhage [62]. Loss of blood volume triggers withdrawal of the parasympathetic nervous system and activation of the sympathetic nervous system which tries to compensate for the drop in blood pressure. As a result, during the early stages of hemorrhage the mean arterial pressure may remain constant and when a significant change in blood pressure is eventually identified, the available medical interventions may be limited. Markov chain methods may describe changes in the compensating autonomic system dynamics related to hemodynamic instability prior to changes in

traditional vital signs, potentially providing an early indicator of hemorrhage.

A Markov chain is defined as a system with different states where the transition probability from one state to the next depends only on the current state, the Markov assumption. A discrete Markov chain can be described by a countable number of states (S) and a transition probability matrix (P) which describes the evolution of a sample path from one state to the other. Regular Markov chains have a limit distribution or steady state. The mixing rate of a Markov chain represents how fast the system is approaching the steady state. Empirical Markov chains can be constructed from sample time series data or first principles. Eigenvalues of Markov chains capture information about changes in system dynamics, which cannot otherwise be captured with nonlinear methods such as Poincare plots [63]. Properties of the transition matrix eigenvalues such as the presence of complex numbers, information content of the limit distribution, and the mixing rate (i.e. the second largest eigenvalue) can be examined to better understand the underlying system.

The change in dynamics of the physiological systems that are represented in the arterial blood pressure (ABP) waveform as the body attempts to compensate for blood loss may be captured by the eigenvalues of the Markov chain. An empirical Markov chain can be constructed from ABP recordings. Each state of the Markov chain can be defined as a blood pressure range (e.g. 80 85 mmHg), and the steady state then represents the probability that the signal is at any one range. As the system dynamics and waveform morphology change, the system will approach steady state faster or slower and this can be observed through the mixing rate.

Here, we present a method to monitor the mixing rate of ABP waveforms. We hypothesized that a detectable change in the Markov chain mixing rate will occur prior to noticeable changes in traditional vital signs in an anesthetized swine model undergoing hemorrhage.

5.3 Our Approach

5.3.1 Experimental Protocol

Experiments were performed at University of Texas Medical Branch at Galveston. The protocol was approved by the University of Texas Medical Branch Institutional Animal Care and Use Committee (IACUC). Immature swine (N=7, female, 37.1 15.1 kg (mean SD)) were propofol anesthetized and instrumented with bilateral catheters in femoral arteries and veins. An arterial pressure catheter was advanced 40 cm into the artery for proximal arterial readings. The carotid artery was catheterized for hemorrhages, a Foley catheter was inserted into the bladder, and a splenectomy was performed. The animal was given a period of at least 30 minutes to recover upon the completion of surgery before data collection.

Data was collected during a slow continuous hemorrhage of 10 ml/kg over 30 min. Physiological monitoring began at least 5 min prior to the initiation of the hemorrhage and occurred throughout the experiment. ABP was recorded using a standard clinical pressure transducer at a sampling rate of 1,000 Hz.

5.3.2 Signal Processing

Heart rate and beat-by-beat blood pressures (systolic, diastolic, mean) were calculated from the ABP waveform using the publicly available code by Zong et al. [106][36]. This algorithm uses a windowed and weighted slope sum function to identify ABP waveform features for each beat. ABP waveform data were down sampled to 125 Hz prior to feature identification. Shock index was calculated by dividing heart rate with systolic blood pressure [82].

5.3.3 Markov Chain Analysis

For computational efficiency, ABP waveforms were down sampled to 100 Hz for the Markov chain analysis. A moving average window (length 2000 samples) was

subtracted from the ABP waveform to remove the effect of a change in the mean arterial pressure from the Markov model.

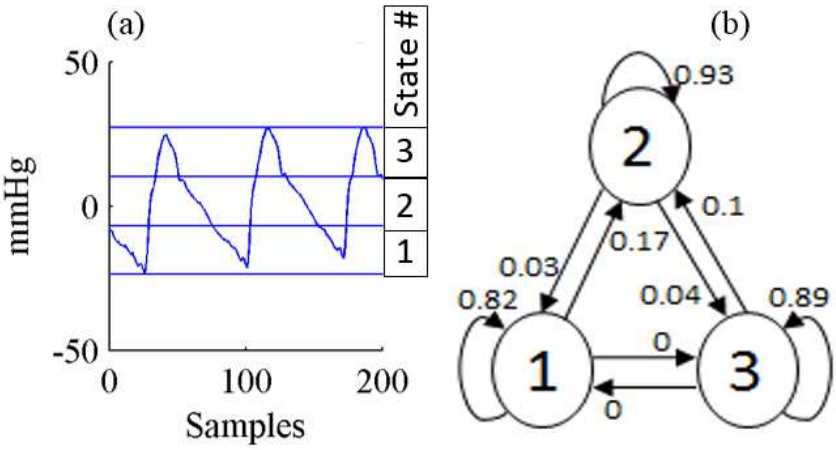


Figure 5.1: Figure 1.(a) 2 seconds recording (sampled at 100 Hz) of an arterial pressure waveform with 3 states. (b) Example Markov chain and its transition probabilities for the three states from the arterial blood pressure wavefrom in (a).

An empirical Markov chain was created from the ABP waveform by segmenting the range of pressure (minimum to maximum pressure recorded for each segment) over a specified window into a fixed number of states, each covering an equal range of blood pressure. Figure 5.1(a) shows a sample ABP waveform filtered using a moving average filter with the same length for two seconds with three states for illustration. The range of blood pressure for each state is computed by dividing the difference of the maximum to the minimum of the blood pressure waveform by three. The empirical transition probability matrix represents the probability that blood pressure will enter any state given only the state that it is currently in. The entry at the ith row and jth column represents the probability with which blood pressure would change from the ith to jth state. To compute the transition probability matrix, the state is defined for each sample by identifying the blood pressure range the sample lies in. Then, the matrix is filled by computing the number of instances a sample moves from state $i(S_i)$ to state $j(S_j)$ over all samples. Finally, the matrix is normalized by dividing

Animal	Heart Rate	Systolic Blood Pressure	Pulse Pressure	Shock Index
A	-0.10	0.47	0.56	-0.59
B	-0.99	0.94	0.98	-0.93
C	-0.09	0.96	0.93	-0.95
D	-0.99	0.98	0.93	-0.98
E	0.36	0.78	-0.31	- 0.82
F	-0.76	0.97	0.96	-0.98
G	-0.98	-0.66	-0.95	- 0.97
Group Statistics				
Median	-0.76	0.94	0.93	-0.95
Min	-0.99	-0.66	-0.95	-0.98
Max	0.36	0.98	0.98	-0.59

Table 5.1: Correlation coefficients between mixing rate and vital signs during hemorrhage.

each row with the sum of the row to have a probability distribution. These probabilities are shown by the arrow labels in the Markov chain in Figure 5.1(b) which corresponds to the example ABP waveform in Figure 5.1(a).

The eigenvalues and the left eigenvectors are determined from the transpose of the transition probability matrix. For a regular Markov chain, all eigenvalues have magnitude less than or equal to 1. 1 is always an eigenvalue, and the eigenvalue with the second largest magnitude is defined as the mixing rate.

We tested a range of window sizes (5 to 30 seconds) and number of states (5 to 30) using the FDA Scientific Computing Laboratory Blue Meadow cluster with Octave parallel computing to identify the appropriate settings for observing changes in the mixing rate of the ABP waveform. The final window length and number of states used for the results reported here were selected as 20 sec and 20, respectively.

5.3.4 Correlation coefficient

Pearson correlation coefficients were determined between the mixing rate and each vital sign (heart rate, pulse pressure and systolic blood pressure).

The mixing rate (determined at 100 Hz of ABP) was first interpolated using a

Animal	Time till Mix-ing Rate changed de-tected (min)	SBP (mmHg)	HR (BPM)	PP (mmHg)	SI (BPM/mmHg)
A	25	-38	0	-11	0.37
B	2	-7	5	-3	0.07
C	7	-23	6	-7	0.35
D	13	-17	21	-13	0.33
E	17	-19	-2	1	0.19
F	33	-19	4	10	0.14
G	12	-1	6	-2	0.04
Group Statistics					
Median	13	-19	5	-3	0.19
Min	2	-38	-2	-13	0.04
Max	33	-1	21	1	37

Table 5.2: Timing of significant change in Mixing Rate and corresponding vital sign changes

cubic spline to match the vital signs (determined at 125 Hz of ABP) that were computed using the Physionet code and all signals were then smoothed using a moving average filter (100 samples window) before computing the correlation coefficients.

5.3.5 Detection of change in the mixing rate

A probabilistic approach was applied to detect a change in the mixing rate during hemorrhage. The distribution during the 5 min baseline period was considered and the 95% confidence interval was determined. A change in the mixing rate was considered at the first instance from the start of hemorrhage when 4 mixing rates (selected from a twelve point window, each of them three points apart) were outside of the 95% confidence interval in the same direction.

5.4 Results and Evaluation

Figure 5.2 shows the vital signs and the mixing rate for each animal starting from 5 minutes before the start of hemorrhage through the 30 min hemorrhage. Hemorrhage was initiated at the red vertical line (0 min). The heart rate exhibited a heterogeneous response to hemorrhage between animals. Heart rate was almost constant in 4 animals while in 3 a dramatic rise in heart rate occurred. Most animals have a steady decline in blood pressure and pulse pressure. Shock index, which captures the changes in both heart rate and blood pressure rises for all animals. The mixing rate decreases consistently in most cases.

The correlation coefficients between the mixing rate and the vital signs quantify the relationship between the two and are presented in Table 5.1. Overall, the mixing rate was inversely correlated with heart rate and positively correlated with blood pressure and pulse pressure, but these were not consistent across all animals. The mixing rate and shock index showed a strong inverse correlation for all animals. The first time that a statistically significant change in the mixing rate was identified for each animal is presented in Table 5.2. There was significant variation between swine in the time that a change was detected, ranging from 2 to 33 min.

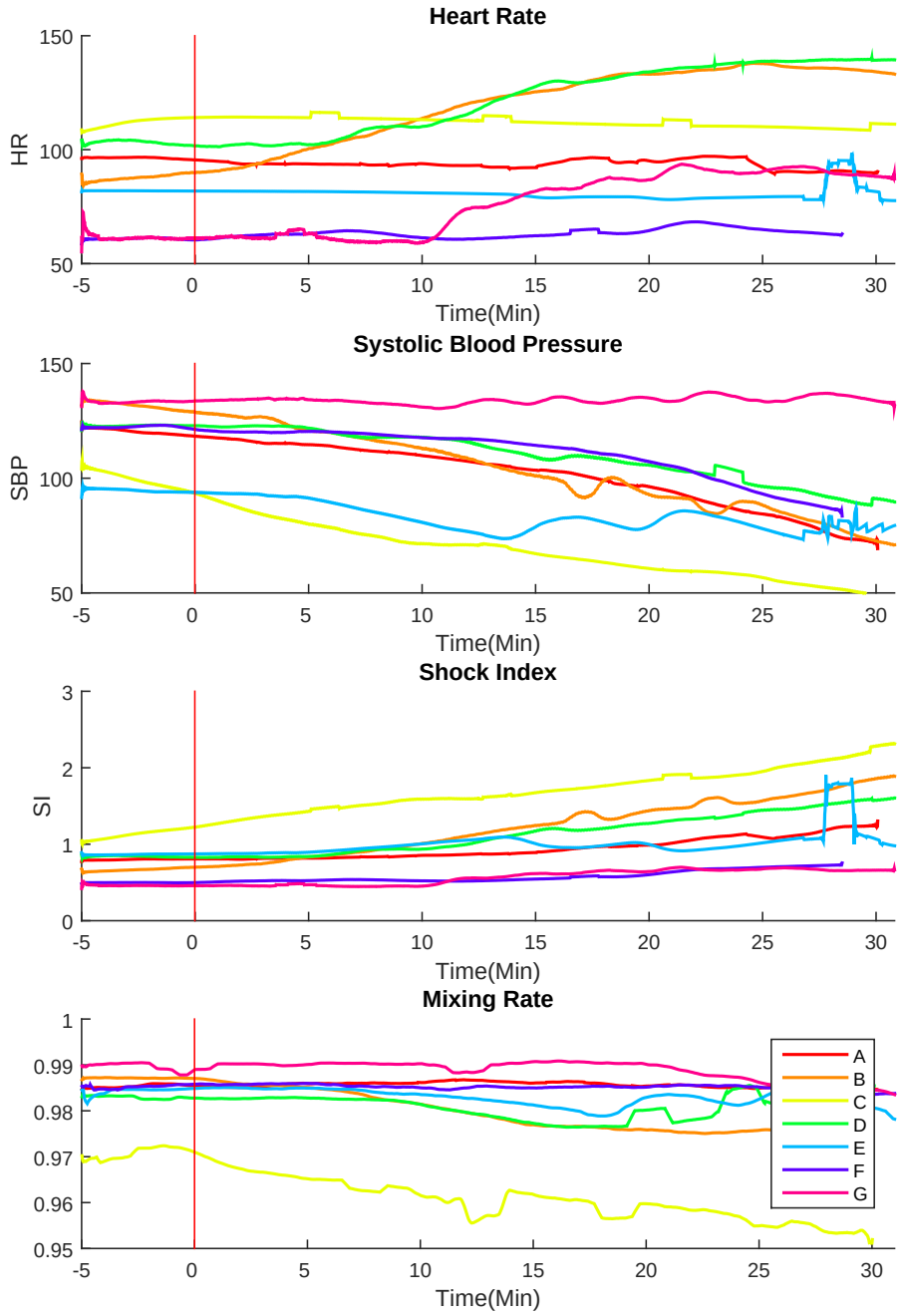


Figure 5.2: The vital signs (heart rate, systolic blood pressure, and shock index) for each animal along with the mixing rate during hemorrhage. Each color represents a different animal.The red vertical lines in each subplot indicate the start of hemorrhage.

The change in systolic blood pressure, heart rate, pulse pressure, and the shock index at the time a change in the mixing rate was detected is also presented in Table 5.2. We see that the mixing rate change is identified for most swine prior to a significant change in the heart rate (median change of 5 BPM), pulse pressure (median change of 3 mmHg) or even shock index (median change of 0.19 BPM/mmHg). At the time a change in the mixing rate was detected, the systolic blood pressure had dropped by a median of 19 mmHg.

5.5 Discussion

In an anesthetized pig model, the Markov chain mixing rate of the ABP waveform is strongly correlated with the vital signs. It has a correlation greater than 0.5 with systolic blood pressure (5 out of 7 animals) and inverse correlation greater than 0.5 with heart rate (4 out of 7 animals) and shock index (7 out of 7 animals).

The relationship between the mixing rate and the shock index indicates that this new variable might be an indicator of impending hemodynamic imbalance. Shock index is widely used in clinical scenarios for identifying patients that need immediate care. It has been shown that patients with an increasing or high shock index have a higher mortality likelihood [21]. The decreasing mixing rate may provide further information about the hemodynamic status of a patient. The shock index is computed from mean vital signs whereas the mixing rate is derived from the dynamics/waveform morphology. As a result, it is not necessary that they provide the same information. The resulting high correlation suggests that the Markov chain mixing rate is capturing changes in the system dynamics or waveform morphology that occur due to the same physiological system changes that affect the mean heart rate and/or blood pressure.

A decreasing mixing rate indicates one special kind of hemodynamic imbalance that occurs due to hemorrhage. Eigenvalues of Markov chains have previously been shown to be identifiers of changes in system dynamics [63]. The construction of the Markov chain from time series data indicates that a change in

waveform morphology or a change in system dynamics might cause a decreasing mixing rate. ABP waveform morphology has been shown to change during central hypovolemia [11]. In the present study, we did not determine if the effects of dynamic or morphological changes in the ABP waveform affect the mixing rate. We suspect that both contribute to the observed decrease in the mixing rate during hemorrhage. To identify if the changes in mixing rate are due to a specific change in waveform morphology, a comparative study of the timing of the changes in the ABP morphology and the timing of the change in mixing rate is needed.

The mixing rate change could be observed for a specific number of states (20) and specific window size (20 seconds). This suggests that the waveform morphology or system dynamics changes can be captured by this method with a 20 seconds waveform segment. This might be useful to understand the utility and the limitation of this method to identify hemorrhage or predicting impending hemodynamic imbalance. Further work is needed to determine the utility of the mixing rate in hemodynamic monitoring.

5.6 Related publications

- **Mohammad Adibuzzaman**, George C. Kramer, Loriano Galeotti et al.: ***The Mixing Rate of the Arterial Blood Pressure Waveform Markov Chain is Correlated with Shock Index During Hemorrhage in Anesthetized Swine*** in Proceedings of EMBC 2014, Chicago, USA

- **Mohammad Adibuzzaman**, Loriano Galeotti, Richard A. Gray, George Kramer, David G. Strauss, Stephen Merrill: ***Markov chain methods in identifying early acute hypotensive episodes*** in MCMi Regulatory Science Symposium, 2013

- **Mohammad Adibuzzaman**, Loriano Galeotti, Richard A. Gray, George Kramer, David G. Strauss, Stephen Merrill: ***Novel physiological monitoring***

strategies for early detection of hemorrhage in Student Poster Session 2013 at the U.S. Food and Drug Administration.

Chapter 6

Heart Rate, Oxygen Saturation, and Perfusion Index: Measuring Vital Signs Using Camera of the Smart Phone

6.1 Introduction

Smart phones with optical sensors have created new opportunities for low cost and remote monitoring of vital signs. In this paper, we present a novel approach to find heart rate, perfusion index and oxygen saturation using the video images captured by the camera of the smart phones with mathematical models. We use a technique called principal component analysis (PCA) to find the band that contain most plethysmographic information. Also, we showed a personalized regression model works best for accurately detecting perfusion index and oxygen saturation. Our model has high accuracy of the physiological parameters compared to the traditional pulse oxymeter. Also, an important relationship between frame rate for image capture, minimum peak to peak distance in the pulse wave form and accuracy has been established. We showed that there is an optimal value for minimum peak to peak distance for detecting heart rate accurately. Moreover, we present the evaluation of our personalized models.

6.2 Related Work

The use of mobile phones for healthcare solutions are creating new opportunities for the solution of one of the greatest problems of humanity, providing healthcare to low income people [52][16]. Studies have been conducted on the use of short message service (SMS) for behavior change [31], the use of questionnaire based symptom assessment system in rural settings [43] and many more. Smart phones that can measure human physiological parameter can bring a rapid change the way healthcare is provided in developing world as well as in developed countries. There

have already a few android and iphone apps that can measure heart rate from facial video as well as from fingertip [1][2].

Christopher et al. [6] used an algorithm depicted in [23] to detect respiratory rate from smart phone video images. Their method used the technique for pulse oxymeter. For oxygen saturation, they used the ratio of the red and blue band and a linear equation to find the correlation. They argument for using red and blue band is that they assumed it would be affected as the same way as the red LED and used blue band as the reference. But there is no valid proof that the red band acts as the same as red LED. They also lack the variation of the constants of the linear equation for different persons. It is unclear from the discussion if this model is for a single subject or it would work for multiple subjects.

For use of green band for analysis of the video image, the authors argued that there is high absorption by hemoglobin in the green range. But there is no mathematical proof that the green band contain the most information. Wim et al. [100] used video image of the facial area for visible light plethysmographic information. The authors argue that due to historical emphasis on PPG signal, visible spectrum has mostly been ignored for plethysmographic information. The authors used a camera with ambient light and the camera was 1.5 meter away from the face of the participants. With this setup, they found up to four harmonics of fundamental HR frequency. They also suggest that not only pulsatile information, but also phase information can be found from the video image. While this experiment is a milestone for visible light photoplethysmographic image, this was the beginning of the use of visible light for pulse oxymetry. The work of Christopher et al. [6]is an extension of [100]. While the authors of [100] found that green band contains the most variability using time frequency transformation, we provide a more robust approach for using green band for frequency analysis.

Ming-Zher et al. [75] used a similar approach of [100] to use facial video

image to extract different physiological parameters namely HR, RR and hear rate variability (HRV) using a webcam. But instead of the green band, they used the band that has the maximum power spectrum of the three bands after independent component analysis (ICA) of the RGB signal of the video image.

The concept of personalized medicine is a relatively new one. According to [3], the Presidents Council on Advisors on Science and Technology states, Personalized Medicine refers to the tailoring of medical treatment to the individual characteristics of each patientto classify individuals into subpopulations that differ in their susceptibility to a particular disease or their response to a specific treatment. Personalized medicine is predominantly an idea from genome technology. Personalized medicine is used for prescribing medication depending upon the molecular structure or genome profiling. It is reducing the traditional trial and error prescribing and making drugs safer [4]. In this context, a personalized model for perfusion index and oxygen saturation is of great interest. The establishment of baseline values for different physiological parameters is also an emerging idea. G. C. Kramer et al. [56] shows that a closed loop control for fluid therapy using multi-parameter model might be of interest. The authors also argue, since the baseline value for different parameters like cardiac output or blood pressure is different for different persons; the baseline values of tissue oxygenation and CO could be encoded in smart tag for a person for quick assessment in injury condition.

6.3 Our Approach

6.3.1 Modeling

The video image was first converted to signals. The mean of each frame is considered one signal and that signal is a time series [Figure 6.1]. Each pixel in the frame has three components; red, green and blue. Most of the work [6][100] done in the literature uses the green band. We first show the reason of using green band by a novel approach by applying principal component analysis which identifies the

component that contribute to the largest possible variance. We found the principal component of the RGB signal or the component corresponding to the largest eigenvalue is very close to the green band. This important analysis validated the use of green band for analysis of the signal. We then used three mathematical model for finding heart rate, perfusion index and saturation of peripheral oxygen (SpO2). For heart rate, we developed a method for finding the optimal minimum peak to peak distance for maximum accuracy given the frame rate or the sampling frequency. For perfusion index and saturation of peripheral oxygen we showed a personalized model would work best to find those parameters accurately using polynomial fit.

6.3.2 Data

We used the camera of a Galaxy Nexus phone with Android 4.1 Jelly Bean operating system. It was used to capture the video image of the index finger for 30 seconds with the flash of the mobile device on. The video image was 1280 times 720 pixels and in 3gp format at a frame rate of 29.5 frames per second. We used Matlab 2012 for all the analysis of the video image. Since Matlab functions only work for avi format, we converted the file to avi using a free converter named Pazera.

6.3.3 Methods

The video image was first converted to signals. The mean of each frame is considered one signal and that signal is a time series [Figure 6.1]. Each pixel in the frame has three components; red, green and blue. Most of the work [6][100] done in the literature uses the green band. We first show the reason of using green band by a novel approach by applying principal component analysis which identifies the component that contribute to the largest possible variance. We found the principal component of the RGB signal or the component corresponding to the largest eigenvalue is very close to the green band. This important analysis validated the use of green band for analysis of the signal. We then used three mathematical model for

finding heart rate, perfusion index and saturation of peripheral oxygen (SpO2). For heart rate, we developed a method for finding the optimal minimum peak to peak distance for maximum accuracy given the frame rate or the sampling frequency. For perfusion index and saturation of peripheral oxygen we showed a personalized model would work best to find those parameters accurately using polynomial fit.

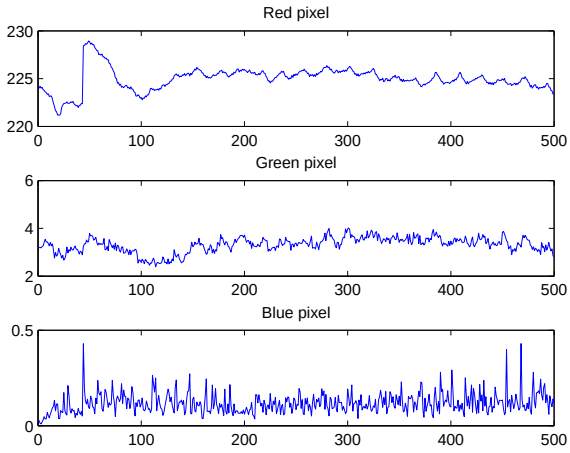


Figure 6.1: Red, Green and Blue component varying from frame to frame. The horizontal axes represent number of frame and vertical axes represent pixel value (between 0 and 255)

Wim et al. [100] and Jonathon et al. [50] used green band to find the heart rate. Each pulse is defined as consisting of the systole and diastole phase of the heart rate. During the systole phase the heart pumps out the oxygenated blood and there is an increase in the blood volume; during the diastole phase there is decrease in the blood volume. We used a technique called principal component analysis (PCA) to find which of the RGB signal is most representative of the signal. Principal component analysis is a very well-known method used for dimensionality reduction. It is used to reduce the dimensionality of the dataset where there is a large number of interrelated variables but retaining the maximum possible variations present in the dataset [49]. PCA finds the signal component corresponding to the eigenvalues in ascending order. The signal

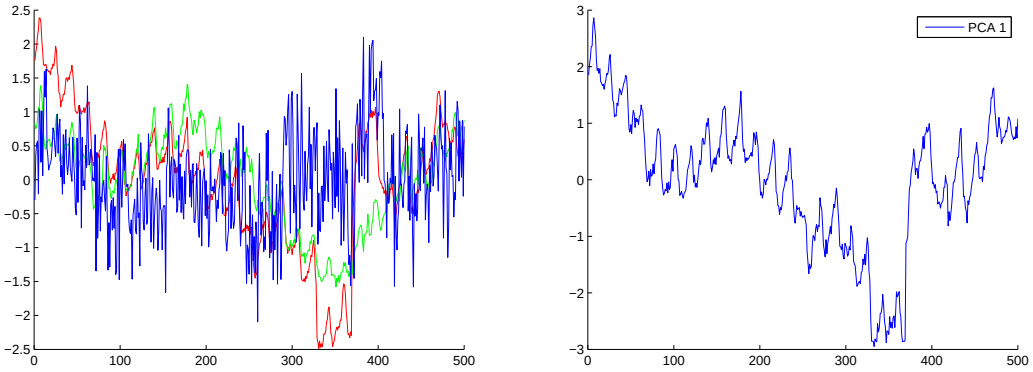
component that has the highest eigenvalue is the most representative of the signal and explains the signal by highest variance. The method uses the offset translated data and create the covariance matrix of the signals. In our case, the red, green and blue signals are highly correlated and that puts the question of which signal to use for finding the physiological parameters. In this context, PCA is a very good fit to find the corresponding signal. We compute covariance matrix, C

$$C = E[B \oplus B]$$

where E is the expectation operator and $\oplus$ is the outer product operator. Then the eigenvalues and eigenvectors of the covariance matrix is computed.

$$V^{-1}CV = D$$

Where D is the diagonal matrix of the eigenvalues. The resulting eigenvalues are then sorted and the source data is converted to the new basis, where eigenvectors form the new basis. We found that the principal component or the projection of the original data corresponding to the maximum eigenvalue is closest to the green band. This similarity is found using Dynamic Time Warping Algorithm. We found the signal corresponding to the maximum eigenvalue is closest to the green band. Figure 6.2 is used to explain this idea. At first the mean of each signal is subtracted from each signal for offset translation and then each signal is divided by the standard deviation for amplitude scaling. The resulting graphs are plotted in Figure 6.2(a) which shows the three components in three different colors. The principal component corresponding to the highest eigenvalue is shown in Figure 6.2(b) which is very similar to the green band.



(a) Red, green and blue pixels using offset translation

(b) Principal component of the red, blue and green pixels

Figure 6.2: Principal component corresponds to the green band.

Heart Rate

Each time the hurt pumps, the red component is increased during the systole phase and is decreased during the diastole phase in the video image. Here, the peak to peak distance represents the time for one heart beat or one pulse. We used this formula to find the heart rate.

$$HeartRate = \frac{(FrameRate * 60)}{NumberofFramesBetweenTwoPeaks}$$

The resulting heart rate was compared to the pulse oxymeter heart rate. For finding the peaks, we needed do define number of minimum frames between two peaks. We found an interesting relationship between frame rate, minimum peak to peak distance and the accuracy. For a particular frame rate, there is an optimal peak to peak distance for finding heart rate accurately. As we increase the minimum peak to peak distance from 4 to 22, the accuracy first increases and then decreases.

Perfusion Index

Perfusion index is defined as the ratio of the pulsatile blood flow to the non-pulsatile static blood flow in a patient's peripheral tissue. Since the systole phase is represented

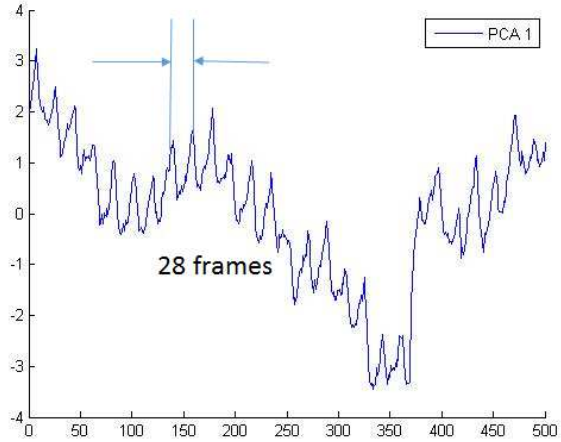


Figure 6.3: Heart rate can be calculated using the formula $HR = \frac{FrameRate*60}{NumberOfFramesBetweenTwoPeaks}$

by the peak of the video image signal and the diastole phase represents the bottom of the video image signal, we used the ratio of each minima to its previous maxima over 30 seconds of time period. Our goal was to find a mathematical model that would map this ratio to the gold standard, which in this case is the pulse oxymeter value of perfusion index.

The non-pulsatile blood flow may also be the average of the red pixel intensity. But we started with the first hypothesis and tried to validate the claim using empirical evidence. We found that for personalized data, that is over time if we take the perfusion index measure for a same person, there is a linear regression equation that explains the perfusion index well with the value of the pulse oxymeter; which is considered as the gold standard for our case. This also explains why a personalized model would work to find the model parameters to detect the perfusion index of a single subject. Figure 6.7 shows the two model with different degrees of polynomial and R^2 value.

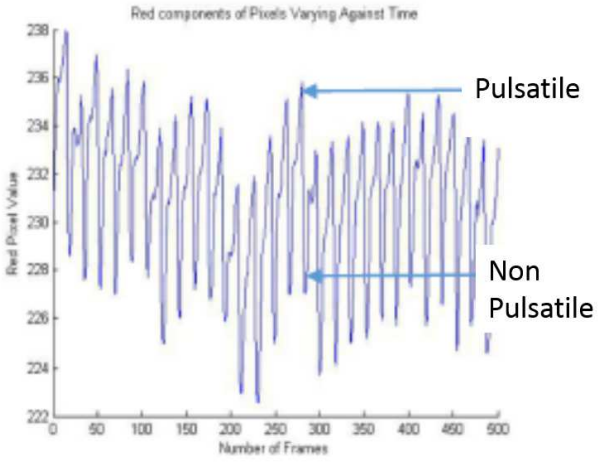


Figure 6.4: Proposed approach for perfusion index which is defined as the pulse strength

Oxygen Saturation

For saturation of peripheral oxygen, it is defined as the ratio of the oxygenated hemoglobin to the ratio of the total hemoglobin.

$$SpO2 = \frac{HbO2}{HbO2 + Hb}$$

To find oxygen saturation we hypothesize to use the peak of each pulse as Oxygenated (because it is bright red and during the systole phase). Then the minima + the maxima is equivalent to HbO2 + Hb For this reason, we used this ratio as an approximation of the oxygen saturation and fit a different degrees of polynomial function to the pulse oxymeter data.

For saturation of peripheral oxygen we have used similar approach as to find perfusion index. We have found that a polynomial fit between the gold standard, which is the pulse oxymeter data for SpO2 and the ratios we have found from the video image requires a higher degree polynomial.

Which in the end leads to an unstable system as the condition number of the coefficient matrix becomes very large. On the contrary, a personalized model with

saturation of peripheral oxygen of a single person can be modeled with even a quadratic equation.

$$SpO2 = \frac{HbO2}{HbO2+Hb}$$

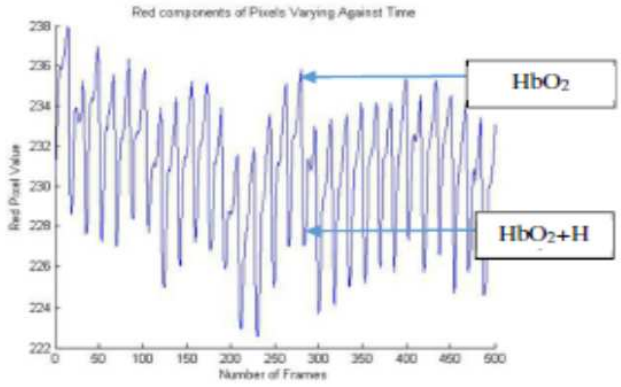


Figure 6.5: Proposed approach for oxygen saturation which is defined as the ration between oxygenated hemoglobin and total hemoglobin.

6.4 Results and Evaluation

We collected video image data for 9 persons. All of them were healthy subjects and between the age of 21 and 32. Seven of them are male and two of them are female. For heart rate, we found a very good accuracy for all the persons. We calculated the perfusion index and oxygen saturation for each of the persons and tried to model the value extracted from the video image using the method discussed in section V. We found that for oxygen saturation and perfusion index, a model takes a higher degree polynomial for a good fit if we try to model the parameters for all the subjects. Figure 6.7 and 6.9 explain this idea. In both of the figure, the X axis represents degree of the polynomial and Y axis represents R^2 value. Interestingly, a model that considers the data for a single subject over multiple time periods, only a quadratic equation gives high R^2 value. This is in connection with our hypothesis that a personalized model for perfusion index and oxygen saturation would be of use in practical purpose. For perfusion index, a simple linear equation can be used for perfusion index. However,

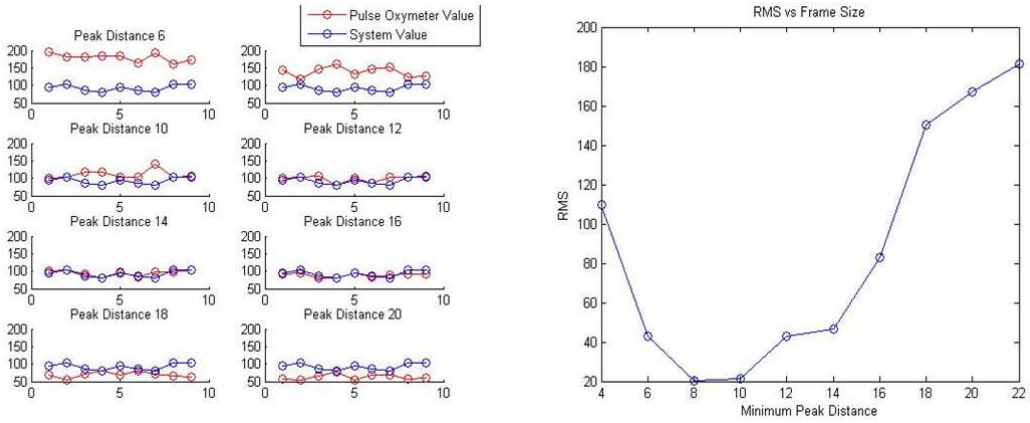


Figure 6.6: It takes a polynomial of degree 5 for polynomial fit of perfusion index for 9 persons with ninety percent data fit. b) Only a linear equation explains ninety percent data for one person.

for oxygen saturation, a linear equation does not give a good fit. But a quadratic equation gives high R^2 value.

6.4.1 Heart Rate

We found for the frame rate of 29.5 fps, the optimal minimum peak to peak distance is 8. The reason behind is that heart rate is varies between 60 and 120 and the sampling rate is the frame rate in this case. If the sampling rate is too high or too low, it skips some of the pulses and that results in inaccurate heart rate. Figure 6.6(a) shows the heart rate data points compared to the pulse oxymeter value for different minimum peak to peak distance. Figure 6.6(b) shows the root mean squared error changes for different minimum peak to peak distances.

This finding is important because it might be possible the frame rates might be different for different cameras. But our model would give the least error by selecting the optimal peak to peak distance based on frame rate.

6.4.2 Perfusion Index

We found that for oxygen saturation and perfusion index, a model takes a higher degree polynomial for a good fit if we try to model the parameters for all the

subjects [Figure 6.7 and 6.9].

In both of the figures, the X axis represents degree of the polynomial and Y axis represents R2 value. Interestingly, a model that considers the data for a single subject over multiple time periods, only a quadratic equation gives high R2 value. This is in connection with our hypothesis

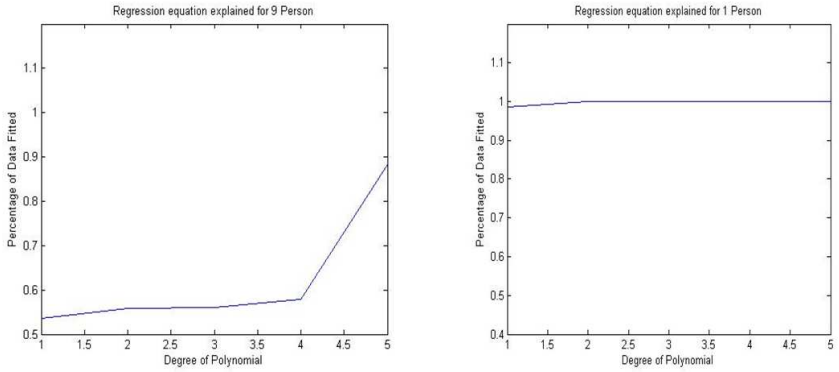


Figure 6.7: a) It takes a polynomial of degree 5 for polynomial fit of perfusion index for 9 persons with ninety percent data fit. b) Only a linear equation explains ninety percent data for one person.

6.4.3 Oxygen Saturation

In general, the app was very close to the readings from the actual device. In some cases, the app gave a reading of 100%, which is not correct. This may have been due to the user turning on the flash. If the flash is on and the finger is covering both the flash and the camera, the readings may be off, as the video will look 100% red from the camera and video processing perspective. Figure 6.8(a) shows a comparison of the readings from the application (red trace) and the actual oximeter device (blue trace). The lines are close, with some measurement error.

Figure 6.8(b) shows a graph of the percent error. The absolute value of the errors consistently falls under 10%, but it is believed that this accuracy can be improved. This will be discussed later in this paper.

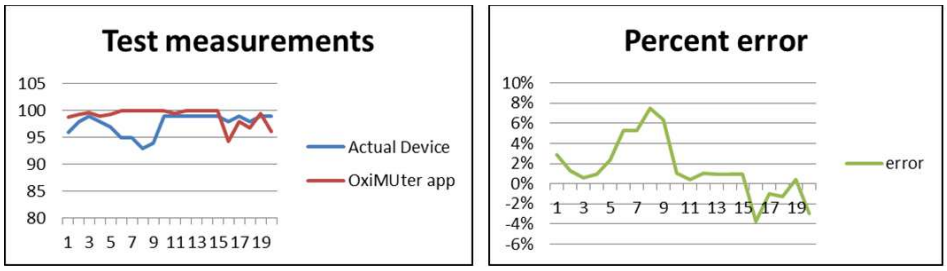


Figure 6.8: Actual error and percentage of error for oxygen saturation.

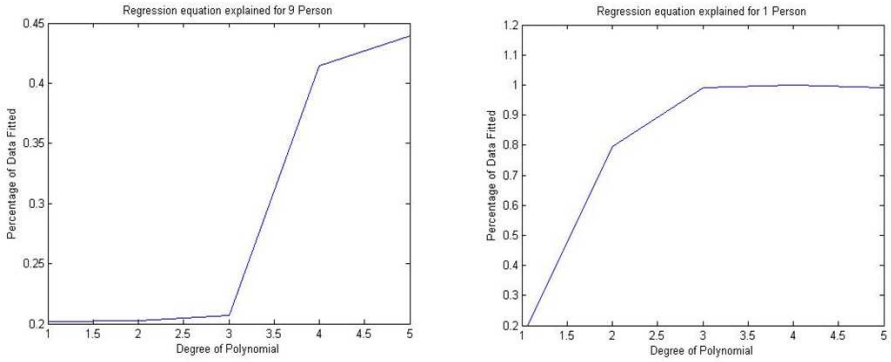


Figure 6.9: a) It takes a polynomial of degree 5 for polynomial fit of perfusion index for 9 persons with forty five percent data fit. b) Only a quadratic equation explains eighty percent data for one person

6.5 Conclusion

The analysis of RGB color space data in time series using principal component analysis is a novel approach for determining the most significant color band. We found PCA output is very close to the green band which validated the use of green band in the literature. Also we showed that for optimal accuracy of the heart rate using video it is necessary to use a minimum peak to peak distance for identifying the systole and diastole phase of the pulse. Use of personalized modeling for perfusion index and oxygen saturation would give a new aspect for these two parameters. Medical science is getting more and more attention in personalized medicine in recent days [14]. One of the critique for your work might be that the features of the signal that are used to model the perfusion index and oxygen saturation may not be the correct one. Our argument in this case is that we used the definition of perfusion index and oxygen

saturation for the model. But work needs to be done to proof the connection of the properties of visible light spectrum to the photoplethysmographic signal.

For regulatory perspective, software applications to measure physiological parameters need FDA approval. But the regulatory organizations still lack the guideline for the personalized measurement of the physiological parameters. But a number of handheld devices and smart phone apps have already been approved by the FDA and the rapid technological innovations are pushing the drive for the regulatory guidelines.

Another major issue is the usability. In our approach, to build the model for perfusion index and oxygen saturation, we need the first few video images labeled with the real value using the gold standard such as traditional pulse oxymeter or laboratory test result. In that context, the application of our system is limited to the patients who regularly visit a healthcare center that offer such solutions. But still this is applicable once the baseline of the parameter for each patient is established. Future physiological parameter baseline could be established using genome profiling.

We showed that the vital signs such as perfusion index and oxygen saturation can have a personalized model. This brings the use of visible light plethysmographic imaging into picture and research in the area of detecting other parameters that can be measured using PPG gets attention. Although historical emphasis on PPG signal blocks the progress for this, the emergence of mobile devices would push the demand for more research on visible light spectrum analysis for photoplethysmographic imaging.

One simple example is finding the arterial fibrillation from video image. Arterial fibrillation is considered as the irregular heart bit due to uncoordinated movements of the arteries. It can be detected using ECG. In our video analysis, we found for some persons, the pulse is not regular at all. It may be the case that those persons have atrial fibrillation due to benign conditions since none of them are

showing any symptoms.

6.6 Related Publications

- **Mohammad Adibuzzaman**, Sheikh Iqbal Ahamed; ”A Personalized Model for Monitoring Vital Signs using Camera of the Smart Phones” in Symposium on Applied Computing, SAC 2013

Chapter 7

Hemoglobin Level: Assessment of Hemoglobin from Mini-video Image Captured by a Mobile Phone

7.1 Introduction

Assessment of hemoglobin levels in human beings is a basic tool in the evaluation of general health and multiple medical conditions. Preliminary studies conducted by our research group show that a model can be developed to achieve high concordance between hemoglobin levels across the usual range in human beings obtained by analyzing mini-video images from a cell phone camera and the usual gold standard laboratory determinations. We conducted a pilot study to investigate the relationship between redness of the blood from a mini-video image of the distal ventral pad of the finger and hemoglobin level as hemoglobin is mostly responsible for the redness of blood. We found significant concordance of the red pixel intensity by mini-video and the hemoglobin level by gold standard venipuncture laboratory testing for a linear least squared regression with a correlation coefficient (r) of 0.68. Red pixel intensity is also influenced by level of oxygenation along with hemoglobin level which in all of our patients was high. In exploring our data we summarized that red pixel color intensity at the fingertip might be influenced by the thickness of the skin which in turn is age and gender-related (men have thicker skin, and thickness decreases with age). We propose to validate our mathematical model for hemoglobin level detection using the camera of a mobile device and develop a healthcare tool for that for effective intervention, improving quality of life and continuous assessment of otherwise unreachable population.

This proposed study is to verify our initial findings for the detection of hemoglobin level from mobile devices from video image of ventral pads of index finger. The application of such system would help improve the quality of life of

patients with chronic diseases such as sickle cell disease, and anemia.

7.2 Related Work

Assessment of hemoglobin levels in human beings is important in evaluation of general health and multiple medical conditions such as anemia, sickle cell disease and many more. Current global practice in medical laboratories involves shining a light through a small volume of blood drawn from patients by venipuncture and using a colorimetric electronic particle counting algorithm to calculate the level. Such demand for laboratory equipment and facilities, and a specific specimen of patient blood obtained with associated discomfort and inconvenience, and the requisite time to obtain results: all make this system less than optimally suited for ideal patient care. The availability of an accurate, rapid and non-invasive means for determining hemoglobin levels which could be used anywhere, would be useful in global medical care.

Prior attempts to develop non-invasive methods of hemoglobin determination have mostly involved hospital monitoring equipment and have given measurement compared to gold standard laboratory assessments with wide variations of -1.7 to +1.8g/dl [61][65][32][69][14][57][35][22]. Gayat et al. recently reported use of a spot device-bedside monitor with a mean error of 0.21g/dl but again the variability of assessment ranged from -3.01 to $+3.42 g/dl$ [34]. Finally an Israeli research group more recently still reported use of a spot device for which they estimated a mean error of $0.1 g/dl$, but still with a variability of -1.59 to $+1.79 g/dl$ [10]. Their device is described as based on occlusion spectroscopy technology in the red/near-infrared range. At the core of this technology is the generation of a new bio-physical signal resulting from temporarily occluding the blood flow in the measurement site. The measurement is performed by using an annular, multi-wavelength probe with pneumatically operated cuffs, which generate an over-systolic pressure at the finger base [40].

In summary, a widely usable, practical, and accurate non invasive method for hemoglobin assessment has not been developed. In our approach for hemoglobin level detection, first, we have read mini-video signals from cell phone camera images of finger pads as red, green and blue light components of varying pixel intensities over time. We have used the average of the red pixel intensity (which should be correlated with red blood cell count), create a calibration table, and determine a calibration constant in a series of patients with different gold standard hemoglobin levels. We found that an analytic method can be defined which can establish high concordance of the redness of fingertip video and gold standard laboratory hemoglobin assessments.

7.3 Our Approach

7.3.1 Experimental Protocol

20 patients from the MCW Emergency Room who are in stable medical condition, and who have otherwise had a hemoglobin determination as part of their clinical evaluations, with these results in hand would, after written informed consent (Appendix I), had a single $30-45$ second min-video taken using a cell phone camera with the flash function on, pressed gently on the ventral bed (the pad) of the tip of the middle finger of the right hand, by the research coordinator. The patients were recruited in 4 blocks of 5 patients each: patients with hemoglobin levels: $>16g/dl$, $13-16g/dl$, $9-13g/dl$, and $<9g/dl$. For the research coordinator, patients were sequentially identified only. The ER physician investigator recorded the laboratory determined hemoglobin level (g/dl) and red blood cell count $(106/ul)$ on the master record, which were later provided to the research coordinator. Figure 7.1 shows the distribution of hemoglobin level for the 17 subjects in the aforementioned four categories. We did not have any patient with $>16g/dl$ hemoglobin level. The age and gender information were also provided after a modification approved by the Institutional Review Board.

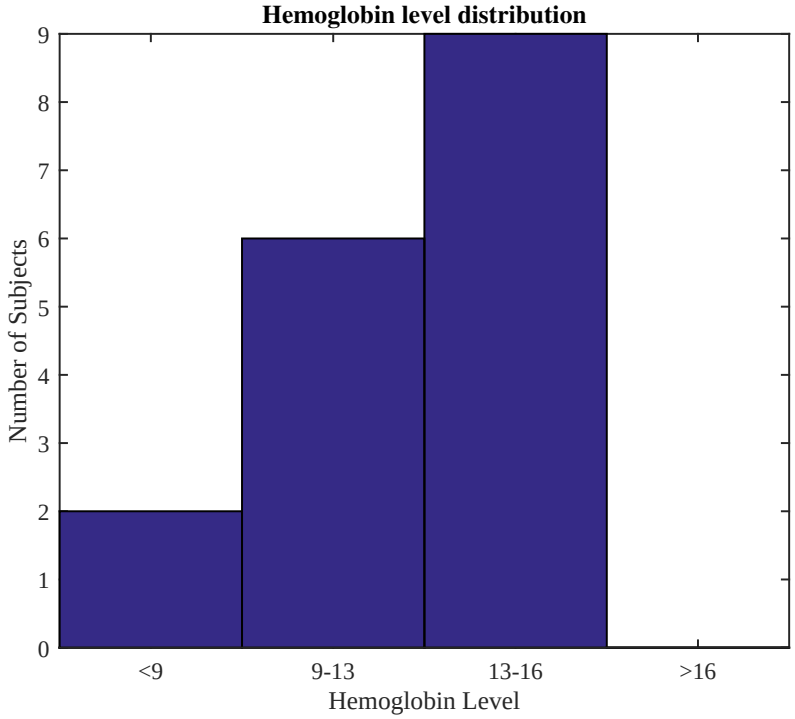


Figure 7.1: Distribution Of Hemoglobin Level

The sequential steps in the study process are listed in Appendix II. Patient eligibility criteria includes

- Hemoglobin level and red blood cell (RBC) count conducted as part of planned appropriate medical care.
- Age $>= 18$ years
- Absence of skin infection involving hands,
- No history of Raynauds phenomenon.
- No administration of greater than 250ml intravenous fluid between time of laboratory hemoglobin assessment sample and the proposed cell phone video assessment.

7.3.2 Data

We used the camera of a Galaxy Nexus phone with Android 4.1 Jelly Bean operating system. It was used to capture the video image of the index finger for 30 seconds with the flash of the mobile device on. The video image was 1280 times 720 pixels and in 3gp format at a frame rate of 29.5 frames per second. We used Matlab 2014 for all the analysis of the video image. Since Matlab functions only work for avi format, we converted the file to avi using a free converter named Pazera.

7.3.3 Methods

Each video image of the finger-tip with the flash of the camera on was converted to three time series data: average of the red, green and blue pixels for each frame over 30 seconds. For each frame, the Y axis represents the average of the red, green and blue pixel intensity of all the pixels in one frame.

For regression analysis, we compared the average of the red pixel intensities over all the frames during a 30 second video and the corresponding hemoglobin level. Figure 7.2 shows the data for the 17 subjects. X axis represents hemoglobin level and Y axis represents red pixel intensities.

7.4 Results

Pearson correlation coefficients were determined between the hemoglobin level and the red pixel intensity over the entire video image. Figure 7.2 shows the data for the 17 subjects. The correlation co-efficient was 0.41 across all the subjects. However, our hypothesis was based on the fact that the skin thickness is similar for all the subjects; and as a consequence, the red pixel intensity in only dependent on the hemoglobin level. But in reality, this is not true. For example, women have thinner skin and also as people age, the skin gets thinner [89][**?**]. As a result, women can show high red pixel intensity even with lower hemoglobin level. To consider the effect of

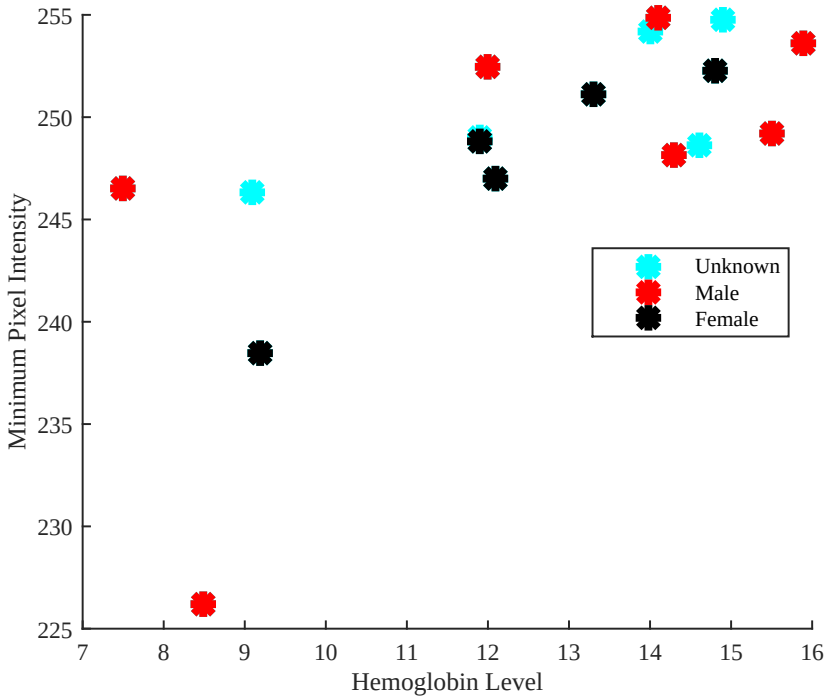


Figure 7.2: Hemoglobin level and red pixel intensity

skin thickness into consideration in our regression model, we hypothesized that the red pixel intensity depends on hemoglobin level as well as age and gender.

$$RPI = \beta_0 Hem + \beta_1 Oxy + \beta_2 ST + \epsilon$$

Where

- RPI =Red Pixel Intensity
- Hem =Hemoglobin Level
- Oxy =Oxygenation
- ST =Skin Thickness
- ϵ = Error

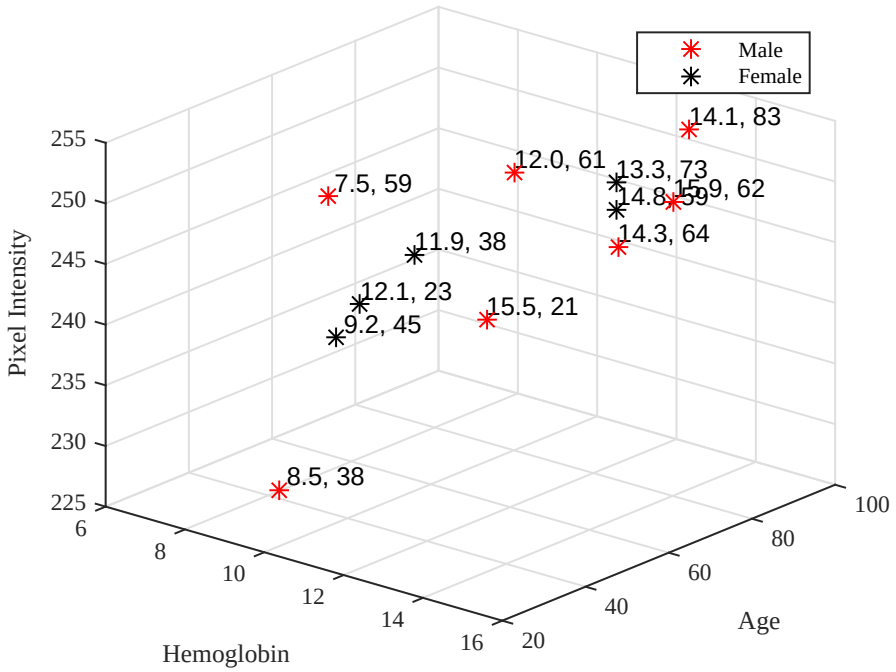


Figure 7.3: Hemoglobin level with pixel intensity, age and gender.

We consider age and gender as corresponding to skin thickness. Further, all the patients had high oxygenation. Figure 7.3 shows the age, gender, hemoglobin level and red pixel intensity. For a linear least squared regression model, the R^2 value is 1, validating our hypothesis.

7.5 Conclusion

To verify our initial finding and to build a robust model for hemoglobin level detection, we need to conduct a large scale pilot study for hemoglobin level detection and consider other factors that can influence the pixel intensities of a finger-tip video image such as pigmentation, skin thickness and oxygenation. Our goal is to collect data for 200 subjects with a varying level of hemoglobin. We would collect data in partnership with our local Non-government Organization (NGO), Amader Gram, in Bangladesh from rural population in Bagerhat, Bangladesh. The skin thickness and oxygenation would also be recorded for the study subjects using ultraviolet skin

thickness measurement device and a pulse oximeter respectively. The data would then be used for a mathematical model development that would map the red pixel intensities of finger-tip video image to the corresponding hemoglobin level. The resulting mathematical model would then be used for creating a mobile phone based healthcare tool for remote and continuous assessment of hemoglobin level with our partner Amader Gram.

Chapter 8

Evaluation: Evaluation of Machine Learning Algorithms For Application in Clinical Setting

8.1 Introduction

In the previous chapters, different mathematical techniques have been discussed for analysis, and monitoring of health parameters for clinical understanding of the patients. For translational clinical research, these algorithms need to be understood in terms of their risk and benefits. There have been much work for algorithmic development for different healthcare problems in recent years. Many algorithms showed promise to advance the healthcare engineering. However, there is much gap between the success of these algorithms in clinical setting and the innovation.

That being the case, there is much need to study the evaluation techniques of these algorithms in a clinical setting that gives us better insights about the risks and benefits of the algorithms. In this chapter, we looked into some of the algorithm development techniques for recommendation of those algorithms in a clinical setting. The influence of the selection of training, testing and validation data set on the performance, namely the sensitivity, specificity and area under the receiver operating characteristics curve which shows the risks and benefits, of different algorithms are investigated.

The study purpose is developing insights on machine learning algorithm development techniques for multi-parameter patient monitoring for the goal of evaluating these algorithms; such that these algorithms can be used for a larger patient population other than the one that has been included in the clinical study. In recent years, machine learning algorithms and the increase of computational power showed promise for mult-parameter patient monitoring using multiple vital signs [46][93][25].

Traditional devices that monitor patients vital signs such as heart rate, blood pressure, oxygen saturation or respiratory rate in the hospital settings have their own individual warning systems for clinical intervention. On the other hand, these devices do not consider the context of the patients underlying physiological status and the relative values of other vital signs. It is argued in medical science and machine learning community that data fusion techniques including machine learning algorithms that consider multiple vital signs would improve early warning systems by reducing number of false alarms (high specificity) and at the same time would represent a better understanding of the patients status and accurately detect or predict critical hospital events (high specificity). Consequently, a good algorithm should have high sensitivity and specificity; which indicates the area under the ROC curve needs to be maximized and the variability of those performance measures should be consistent across random selection of the target patient population. Questions remain how to evaluate these algorithms, such as

- How to select the training, testing and validation data set for developing the algorithms?
- Do the algorithms outperform the existing reference system for early warning score?
- And, if the algorithms have less number of false alarms and thus help solve the problem of alarm fatigue.

8.2 Our Approach

A large clinical data set was used for random selection from the target population for the study. The algorithm development technique is mimicked from real life scenarios and was divided in three stages: 1) training, 2)testing and 3) validation. The independent validation set was pre-selected and was same for the validation phase.

The training and testing data sets were based on different sizes and different random splits.

These data sets were used for training and testing, and the algorithms were designed to predict clinical events. One such clinical event is called critical hospital event, for which urgent care is needed. Medical Emergency Activation Criteria (MET Activation Criteria) is considered as one such critical hospital event; upon the activation of these events, a team consisting of medics are activated for urgent care of the patient. The use of such medical emergency team has been proved to improve patient outcome [13].

8.2.1 Data

For our study, we used a large publicly available data set in intensive care unit of Beth Israel Deaconess Medical Center called Multi-parameter Intelligent Monitoring in Intensive Care (MIMIC II). The database containts high resolution temporal data including lab results, electronic documentation, bedside monitor trend data and waveforms. The data set has two components: clinical database and waveform database. They also have a matched subset of 6000 records that matches the clinical database record to the corresponding waveform database record (Figure 8.1).

The clinical database contains de-identified information regarding the clinical care of patients as well as data from hospital archives; including lab tests, hourly vital signs, ventilator settings, fluid intakes etc. The waveform database contains high resolution vital signs including electrocardiogram data, blood pressure, oxygen saturation and respiratory rate.

2400 records remained after applying the inclusion exclusion criteria. Inclusion exclusion criteria includes, a) the records are matched across clinical data set and waveform data set, b) at least 5 hours of continuous trend data for heart rate, oxygen saturation, systolic blood pressure and respiratory rate were present (Figure 8.2). Of these 2400 records, 400 records were selected as an independent validation set

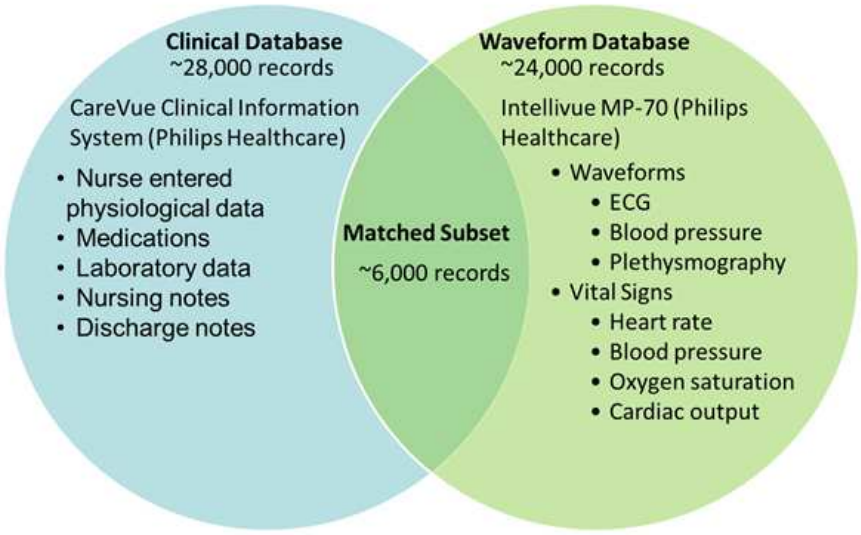


Figure 8.1: Components of Multi-parameter Intelligent Monitoring in Intensive Care (MIMIC II) data set. Image Courtesy: Scully, C.

and 2000 records were used for training and testing. These 2000 records were randomly split into different subsets for creating different training sizes. For example there were 1 subset of size 2000, 1 subset of size 1500, 2 subsets of size 1000 and 4 subsets of size 500. For each subset, the training, testing and validation was performed independently. The algorithm development for each subset was done by a 10 fold cross validation with 10 different random splits for training and testing. The best algorithm for each subset was selected as the one from the cross validation that gives the best accuracy using the test sets.

8.2.2 Study Design

A moving window of 130 minutes that shifts each minute at a time was selected. Of these 130 minutes, the first 60 minutes was defined as the observation window, the next 60 minutes were defined as the gap window and the last 10 minutes were set as the target window (Figure 8.3). The goal was to predict a critical hospital event in the 10 minute target window, 60 minutes in advance, which is defined as the gap window. The prediction algorithm were fed the vital signs data of 60 minutes

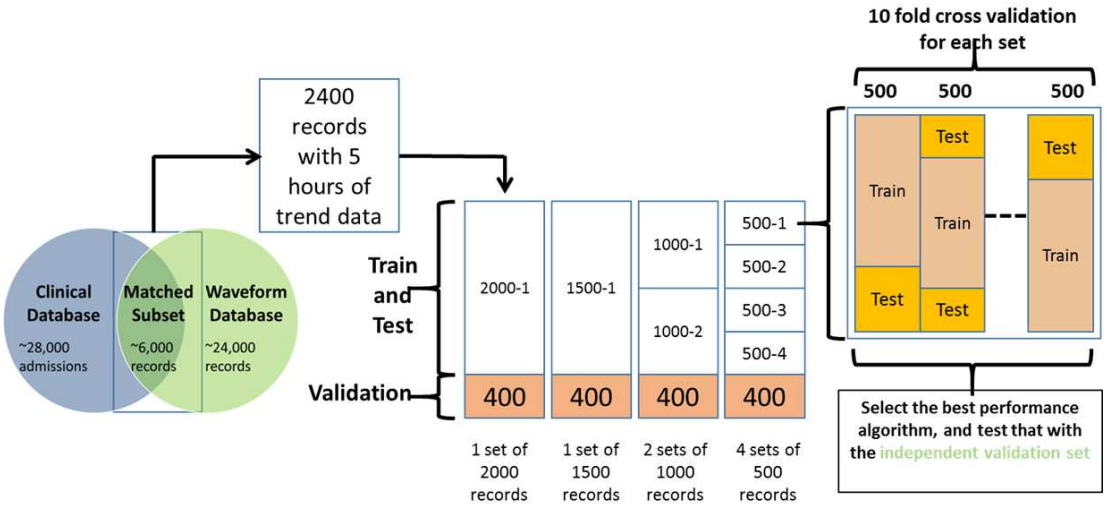


Figure 8.2: Selection of training, testing and validation data set for the study.

before the gap window. This window of 60 minutes, for which the machine learning algorithms used different features such as mean, variance, kurtosis of the vital signs is defined as the observation window. Figure 8.3 shows how the observation window, gap window and target window was defined for the study.

8.2.3 Algorithms

For the study, two machine learning algorithms were used: decision tree and support vector machine (SVM). For each of the algorithms, 19 features were selected and fed into the machine learning algorithms.

Features include mean, median, variance, kurtosis, skewness from each of the vital signs and the cross correlation coefficient between any two vital signs. Additionally, the early warning score using the National Early Warning Score was selected as another feature [5]. Figure 8.4 shows the feature selection and the flow of the algorithms for the prediction of medical emergency team activation.

Medical emergency team (MET) activation criteria

Medical emergency team (MET) activation criteria is used in hospital settings as clinically critical events and was set as the event to predict in our study. The MET

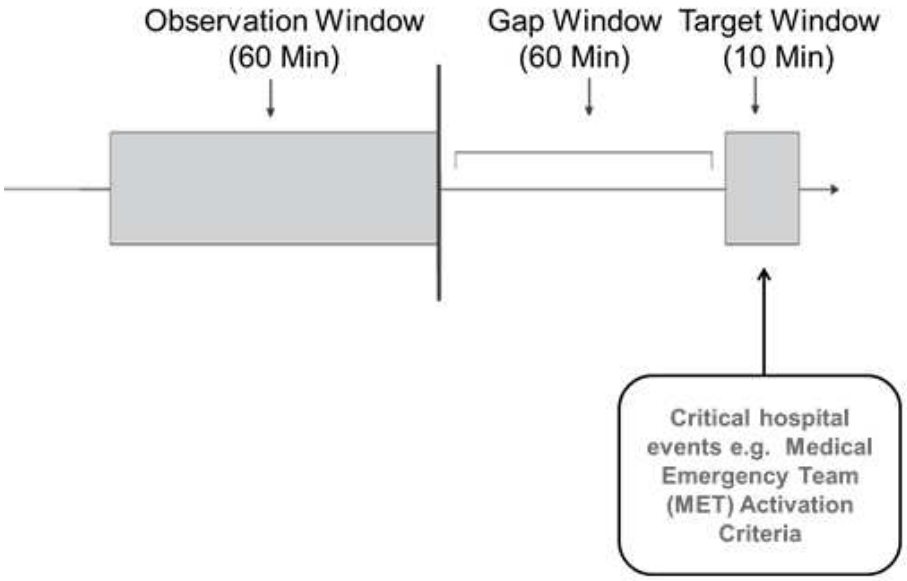


Figure 8.3: Observation, gap and target window for the prediction of critical hospital events from vital signs.

criteria for the three vital signs: heart rate, systolic blood pressure and oxygen saturation is given below. If one or more criteria is satisfied in the target window, a true positive event was considered. MET activation criteria includes:

- Heart rate (< 40 or > 160 bpm)
- Systolic blood pressure(< 60 or > 200 mmHg) or
- Oxygen saturation ($< 85\%$)

The goal of the algorithms is to predict if MET activation will occur 60 minutes in advance (in the target window)

For example, Figure 8.5 shows a true positive event as the oxygen saturation drops below 85% for at least 9 minutes in the target window of 10 minutes. Figure 8.6 shows the example of a true negative window, in which case all the vital signs are within the normal range in the target window.

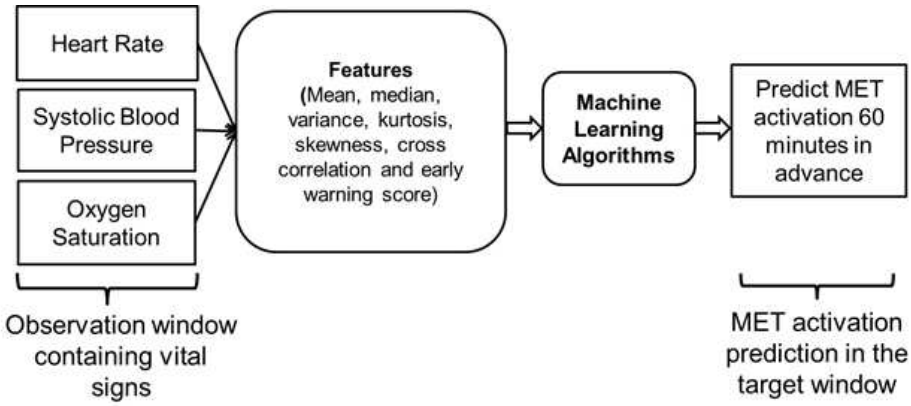


Figure 8.4: Feature selection and use of machine learning algorithms in the study.

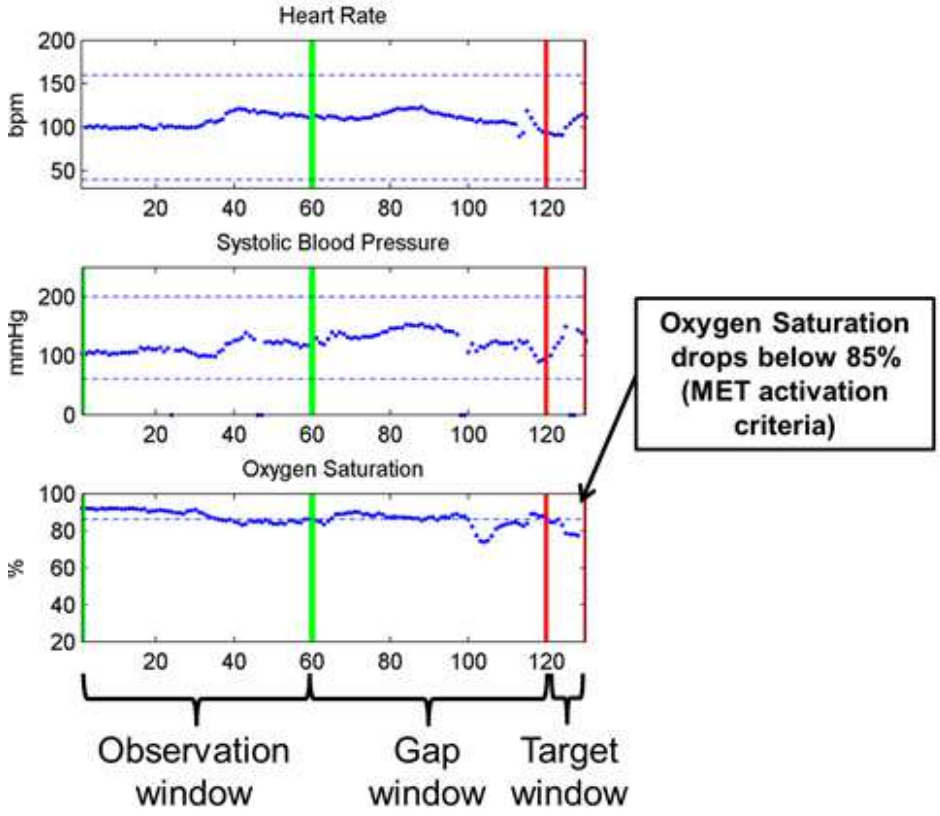


Figure 8.5: Example of a 'true positive event' using the MET activation criteria. The green vertical lines denote the observation window and the red vertical lines denote the target window.

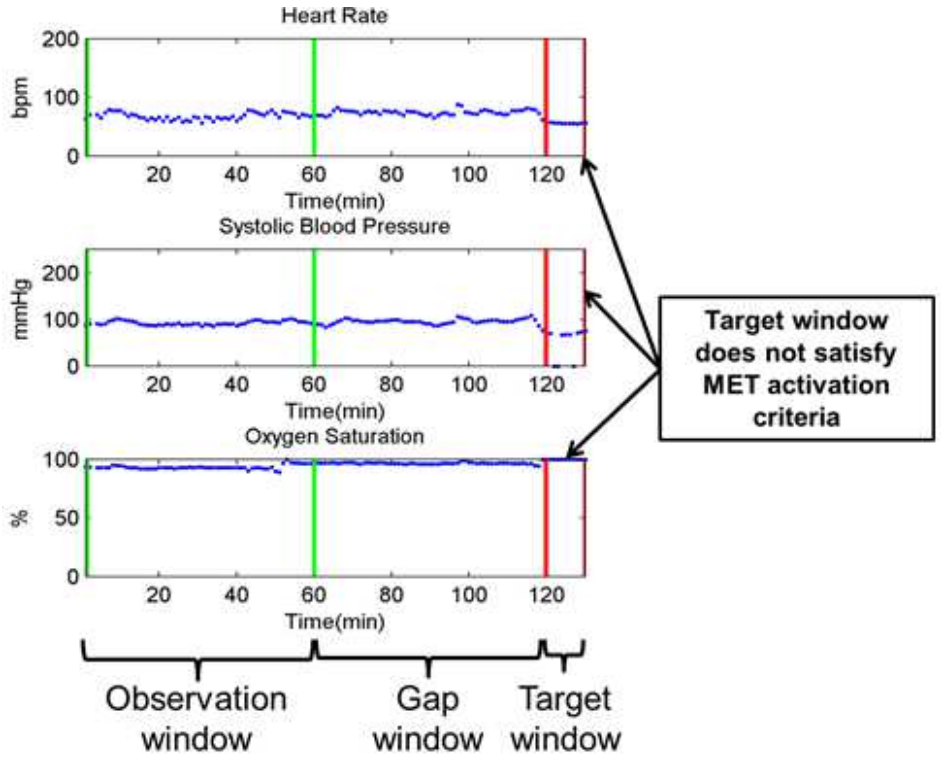


Figure 8.6: Example of a true negative window for the MET activation as the target window does not satisfy MET activation criteria.

Decision Tree

Decision tree is a machine learning algorithm for binary classifier. For classification of true positives and true negatives from the training set, it looks into all the features and all the thresholds such that it can maximally discriminate between true positives and true negatives. That threshold is set as the root. If one threshold cannot discriminate all the true positives and the true negatives in the training set, it iterates over all the features and the thresholds again and creates a multilevel decision tree as the classifier. The thresholds on each node depend on the training set. Figure 8.7 shows a sample decision tree for medical emergency team activation.

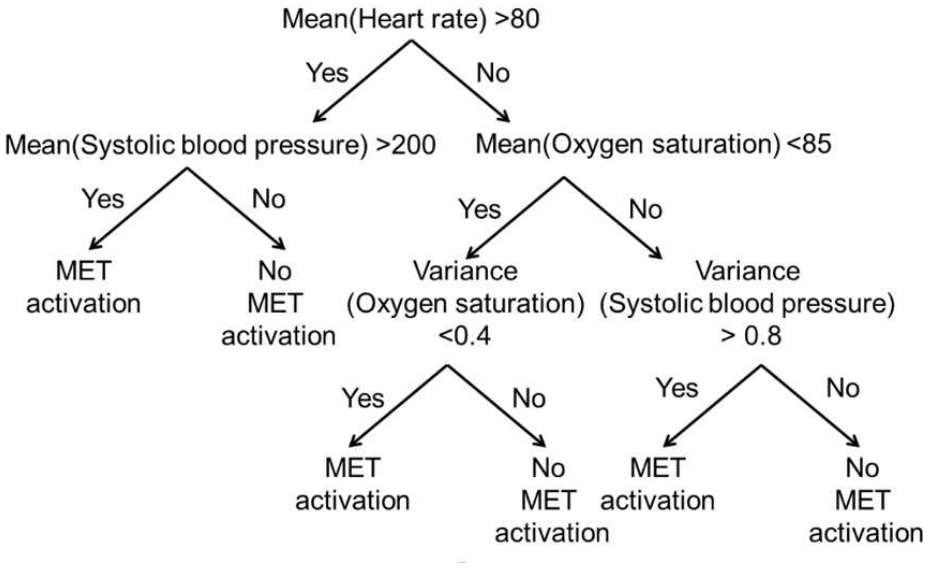


Figure 8.7: Example of a decision tree for MET activation.

Support Vector Machine (SVM)

Support vector machine is another binary classifier that classifies the true positive events and the true negative events using a hyperplane. The hyperplane is found by using the classifier function,

$$y(x) = W^T \psi(x) + b$$

Where ψ is a continuous feature space transformation, W is the weight vector and b is the bias parameter. Figure 8.8 shows the basic concept of support vector machine using a simple two dimensional feature space.

National Early Warning Score (NEWS)

For our study, we used the early warning score recommended by the review committee at the Royal College of Physicians for identification of future patient deterioration. This was criteria was considered as the 'gold standard' against the algorithms. Figure 8.9 shows the thresholds for different vital signs for different scores of NEWS.

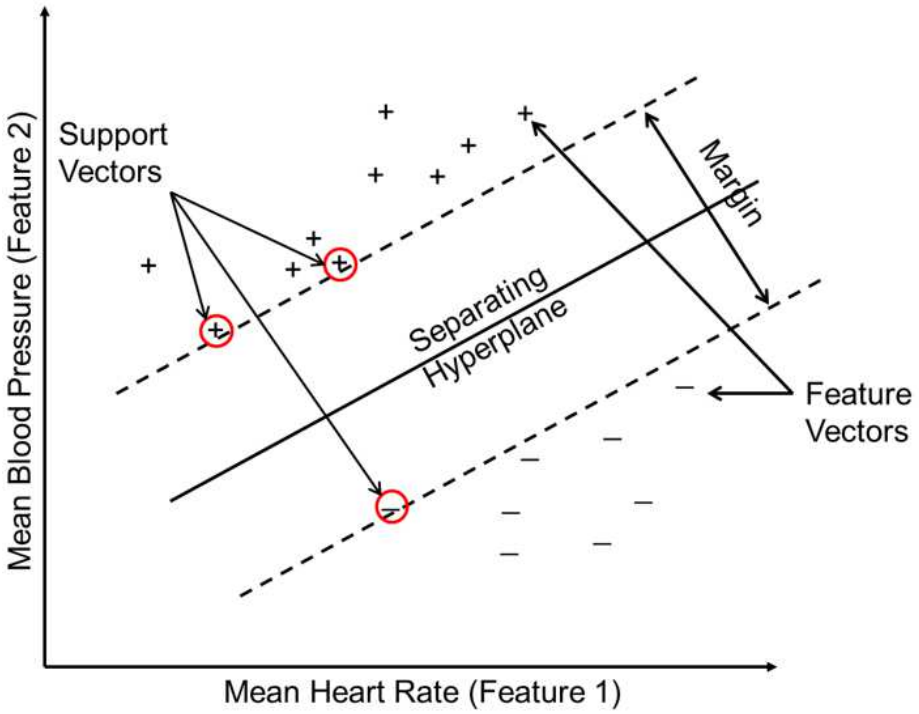


Figure 8.8: Support vector machine for binary classification using two dimensional feature space.

PHYSIOLOGICAL PARAMETER	3	2	1	0	1	2	3
Respiration Rate	≤8		9-11	12-20		21-24	≥25
Oxygen Saturation	≤91	92-93	94-95	≥96			
Systolic BP	≤90	91-100	101-110	111-219			≥220
Heart Rate	≤40		41-50	51-90	91-110	111-130	≥131

Figure 8.9: Thresholds for different vital signs for different score using NEWS.

A normal range for the vital signs are scored as 0. As the vital signs deviates from the normal range, lower or higher, the score is higher (between 1 and 3). A true positive event (prediction for MET activation) using NEWS was defined,

- If any of the vital signs reaches the threshold of 3 for 10 minutes in a 60 minute observation window.
- If the cumulative score reaches 5 for 10 minutes in a 60 minute observation

window.

8.3 Results

The results were analyzed for the three approaches: two machine learning algorithms, and the reference system in terms of sensitivity and specificity. Figure 8.10 shows the sensitivity and Figure 8.11 shows the specificity for the three approaches.

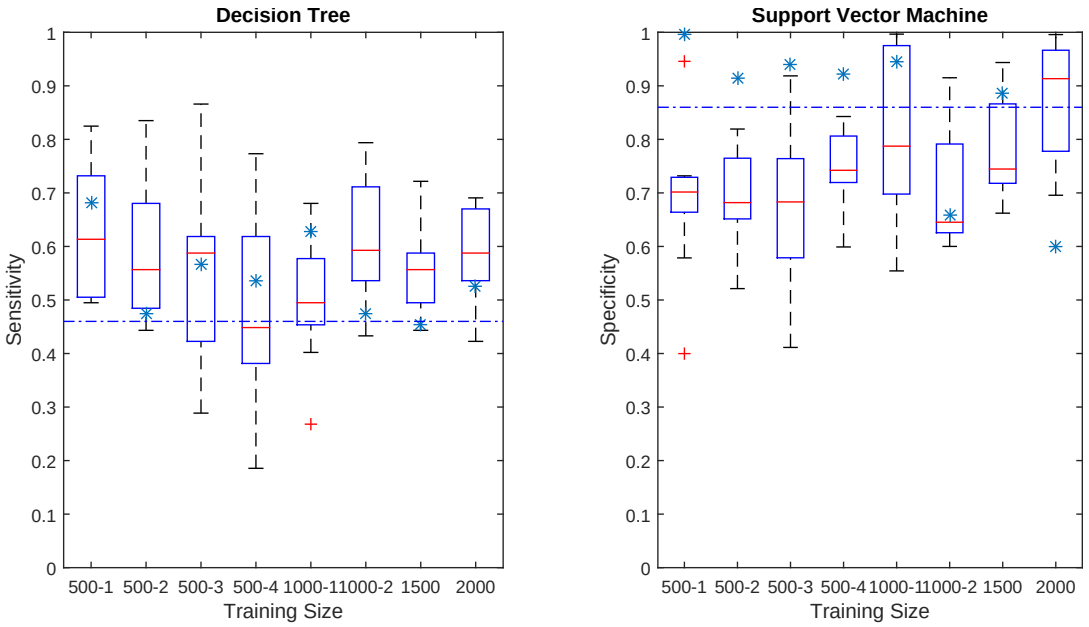


Figure 8.10: Sensitivity for the machine learning algorithms for different training sizes. X-axis represents training size and trial. For each training set, the results of a 10 fold cross validation are reported as box plots (the central red line is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the extreme data points the algorithm considers to be not outliers, and the red + sign denotes outliers). The blue asterisks represent the performance on the validation set of the algorithm that performs best on the test set. The blue dashed lines represent the performance of NEWS score.

The results indicate that increasing the training size does not necessarily increases the performance of the algorithms. Also, the increase in performance relative to the increase in the training data size is algorithm dependent. For example, for lower training size, decision tree has higher sensitivity, but lower specificity; but it is the other way in case of support vector machine.

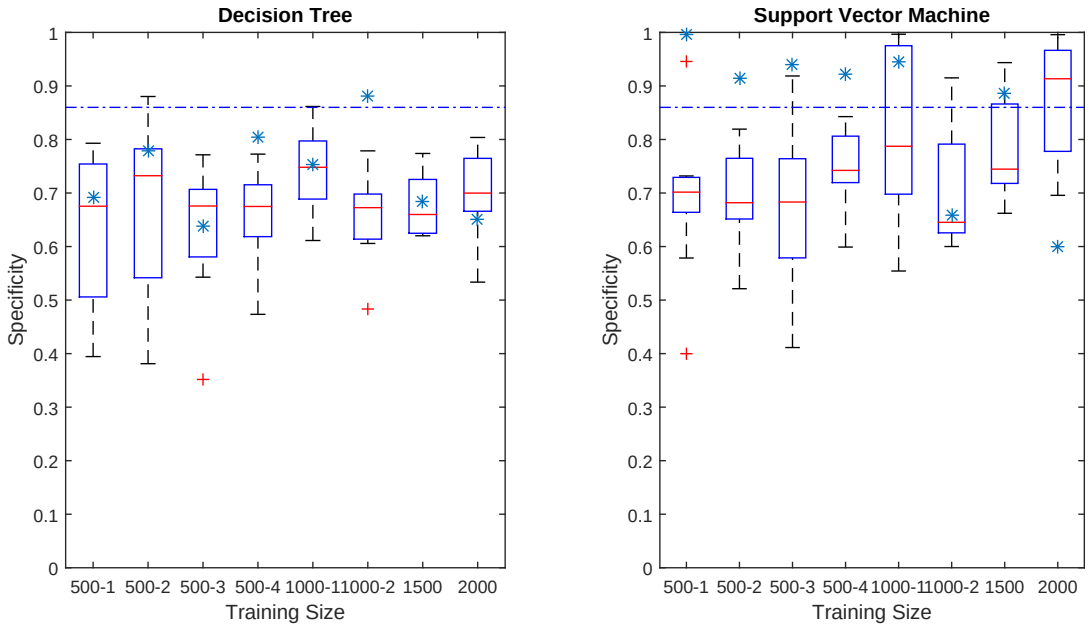


Figure 8.11: Specificity for the machine learning algorithms for different training sizes. X axis represents training size and trial. For each training set, the results of a 10 fold cross validation are reported as box plots (the central red line is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the extreme data points the algorithm considers to be not outliers, and the red + sign denotes outliers). The blue asterisks represent the performance on the validation set of the algorithm that performs best on the test set. The blue dashed lines represent the performance of NEWS score.

Furthermore, the best algorithm in the testing phase, does not necessarily performs best for the independent validation set. Often, the best algorithm for the independent validation set does not represent the best algorithm in the training phase using the test set. This variability of the performance of the algorithms should be minimized for a robust machine learning algorithm.

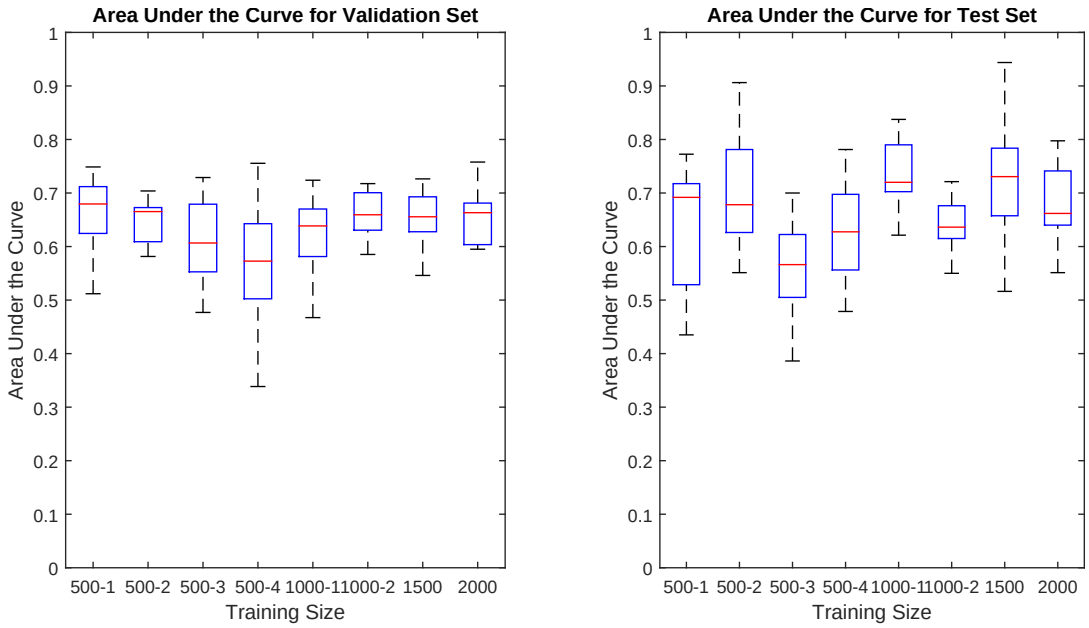


Figure 8.12: Area Under the Curve (AUC) for different Receiver Operating Characteristics Curve (ROC) for different thresholds using decision tree. X axis represents training size and trial. For each training set, the results of a 10 fold cross validation are reported as box plots (the central red line is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the extreme data points the algorithm considers to be not outliers, and the red + sign denotes outliers).

To understand the variability of the machine learning algorithms across cross validation and between test phase and validation phase, the area under the curve for different cross validations, different training sizes are plotted in Figure 8.12. Here, it is found that there is much variability within the same training size in terms of the area under the curve (AUC) of the receiver operating characteristics. But there are instances when the AUC is much higher for the test test than in validation set, which may lead to the 'false belief' that an algorithm has good performance using the test set, and eventually be a high risk algorithm in a clinical application with an independent validation set.

8.4 Discussion

We could not reach any definite conclusion as to how we can claim the best machine learning algorithm for a clinical application. However, we have identified there remains significant risk in defining an algorithm as 'high performing' without an independent validation set. In addition to that, increasing the training size and the performance increase in the algorithms is very algorithm dependent. It usually depends on the size of the hypothesis space for that particular algorithm. In general, a good algorithm should have less variability in the performance across cross validation and needs to be tested with an independent validation data set for minimizing the risk in a clinical setting; as a lower performance of these algorithms could be life threatening. But as the algorithms show prospect for performing the existing system, a 'good' algorithm has the potential to be beneficial in clinical setting and improve patient outcome.

8.4.1 Challenges

Key challenges to build such systems include:

- How to deal with large amount of data.
- How to define a clinically meaningful event without any annotation in the data set and
- How to compare the results without any existing reference systems.

To address the first challenge, we used the FDA supercomputer (Betsy Cluster) to split the data set and to search into the data set for true positive and true negative events with octave parallel computing and 400 nodes. In the case of the second challenge, we build our system with the definitions from physiology (Medical Emergency Activation Criteria) that define a true positive event, a true negative event. We also used the current methods used by the physicians to address the third

challenge, by using the National Early Warning Score (NEWS) as the reference system. We argue that such algorithms should use domain specific knowledge for the definitions of true events and improve the performance in that particular domain.

8.4.2 Future Work

The study design for this work is not perfect and there are different issues that needs to be addressed. The definition of a true positive event and a true negative event is important and must represent the underlying physiological condition of the patient. There are many ways that could represent a false situation of the patient because of noise, displacement of the sensor, movement of the subject, and different baseline for different individuals. In our study design, these factors are not considered.

Not to mention, the algorithm development phase needs to be studied carefully. We have selected 19 features for the study. They are mean, median, variance, kurtosis, skewness from each of the vital signs and the cross correlation coefficient between any two vital signs. Additionally, the early warning score using the National Early Warning Score was selected as another feature. But all of these features does not differentiate the true positive event and the true negative events significantly. Only those features should be selected, that differentiate that distinguishes the positive and negative classes adequately.

Besides, the number of false alarms in this particular setting needs to be considered as another performance measure for multi-parameter patient monitoring along with the receiver operating characteristics or area under the curve. The end goal is to have less false alarms with high sensitivity.

Alternative techniques and the mathematics of the algorithms also needs to designed carefully as to represent the clinical situation. For example, a decision tree algorithm might be a good fit for this setup, but a neural network may not; as the patient monitoring phase works as different levels of a decision, similar to the decision tree algorithm.

In summary, future works to improve early warning patient monitoring algorithms should investigate:

- Careful definition of true positive and true negative events.
- Number of false alarms.
- Key signal features to include in the algorithms.
- Alternative techniques to combine information from multiple sources.

Chapter 9

Conclusion

9.1 Summary

We have built systems, developed algorithms for monitoring of health parameters and explored the evaluation of these algorithms in a clinical setting. Despite much advance in hardware and software as well as computational power during the last decade, there is much gap between the advancement in technologies and the application of those technologies in a clinical setting. The research work that utilizes this advanced sensors is presented here so that it can have an impact by addressing the issues of system deployment, utilizing the mobile sensors, and evaluation techniques of the algorithms in a clinical setting. There is much scope for future work to advance the knowledge in this area.

For affect detection from multiple modalities, other modalities can be incorporated in the model such as heart rate, skin temperature and pupil size. For pain level detection, the images can be pre-processed for adjustment of lighting condition and face extraction. For hemoglobin level detection, a large pilot study is being undertaken in Bangladesh to validate the model. For evaluation approaches, careful construction of the study is needed to look at the results in terms of number of false alarms and defining the true positive and true negative events.

9.2 Contributions

The contribution of the research work is in algorithm development, system design, usability for monitoring of health parameters and evaluation of the algorithms for application in a clinical setting. The contribution of the work is discussed in relationship to each of the applications.

9.2.1 Affect: Smart Phone based Affect Detection

Most of the research in multi-modal affect detection has been done in laboratory environment. Little work has been done for *in-situ* affect detection. In this study, affect detection in natural environment using sensors available in smart phones was explored. Facial expression and energy expenditure of a person were used to classify a person's affective state by continuously capturing fine grained accelerometer data for energy and camera image for facial expression and measure the performance of the system. The system was deployed in natural environment and was provided special attention on annotation for the training data validating the 'ground truth'. Important correlation between facial image and energy was found, which validates Russell's two dimensional theory of emotion using arousal and valence space.

9.2.2 Pain Level: Smart Phone Based Pain Level Detection

The hypothesis of using smart phone cameras for pain assessment of breast cancer patients in a developing country is validated in this study. A smart phone based system was built that detects pain intensity from facial images using the Eigenface method which is based on principal component analysis (PCA). The weight vectors of the Eigenfaces are classified using a Euclidian distance, angular distance and support vector machine (SVM). The system performs best with a support vector machine when the performance measure is the distribution of the input and output labels. The system uses facial images submitted by rural breast cancer patients in Bangladesh, which were taken with the camera of a mobile phone. Each image was assigned a pain value by the users. The data were collected as a part of a six month long longitudinal study funded by International Breast Cancer Research Foundation (IBCRF). We show that a personalized model for pain assessment works better for automatic pain assessment in this setting. This work is a proof of concept of using smart phones for remote pain assessment which might be a game changer to provide low cost health care solutions

while the labeled pain image data set could be used to identify real world insights for community reuse.

9.2.3 Arterial Blood Pressure: Novel Index for Identifying Early Markers of Hemorrhage

Identifying the need for interventions during hemorrhage is complicated due to physiological compensation mechanisms that can stabilize vital signs until a significant amount of blood loss. Because the physiological systems providing compensation affect the arterial blood pressure waveform through changes in dynamics and waveform morphology, Markov chain analysis of the arterial blood pressure waveform was used to monitor physiological system changes during hemorrhage. Continuous arterial blood pressure recordings were made on anesthetized swine (N=7) during a 5 min baseline period and during a slow hemorrhage (10 ml/kg over 30 min). Markov chain analysis was applied to 20 sec arterial blood pressure waveform segments with a sliding window. 20 ranges of arterial blood pressure were defined as states and empirical transition probability matrices were determined for each 20 sec segment. The mixing rate (2nd largest eigenvalue of the transition probability matrix) was determined for all segments. A change in the mixing rate from baseline estimates was identified during hemorrhage for each animal (median time of 13 min, 10% estimated blood volume, with minimum and maximum times of 2 and 33 min, respectively). The mixing rate was found to have an inverse correlation with shock index for all 7 animals (median correlation coefficient of -0.95 with minimum and maximum of -0.98 and -0.58, respectively). The Markov chain mixing rate of arterial blood pressure recordings is a novel potential biomarker for monitoring and understanding physiological systems during hemorrhage.

9.2.4 Heart Rate, Oxygen Saturation, and Perfusion Index: Measuring Vital Signs Using Camera of the Smart Phone

In this chapter, we present a novel approach to find heart rate, perfusion index, oxygen saturation and hemoglobin level using the video images captured by the camera of the smart phones with mathematical models. Principal component analysis (PCA) was used to find the band that contain most plethysmographic information. Also, it was shown that a personalized regression model works best for accurately detecting perfusion index and oxygen saturation. The model has high accuracy of the physiological parameters compared to the traditional pulse oxymeter. Also, an important relationship between frame rate for image capture, minimum peak to peak distance in the pulse wave form and accuracy has been established. It was showed that there is an optimal value for minimum peak to peak distance for detecting heart rate accurately.

9.2.5 Hemoglobin Level: Assessment of Hemoglobin from mini-video image captured by a mobile phone

Assessment of hemoglobin levels in human beings is a basic tool in evaluation of general health and multiple medical conditions. In current global practice this is done in medical laboratories by shining light through a small volume of blood drawn from patients by venipuncture and using a colorimetric electronic particle counting algorithm to calculate the level. Such demand for laboratory equipment and facilities, and a specific specimen of patient blood obtained with associated discomfort and inconvenience, and the requisite time to obtain results all make this system less than optimally suited to ideal patient care. This study shows that a method can be developed to achieve high concordance between hemoglobin levels across the usual range in human beings obtained by analysis of mini-video images from a cell phone camera and the usual gold standard laboratory determinations. We conducted a pilot

study to investigate the relationship between red pixel intensity from a mini-video image of the distal ventral pad of the finger. Under an ethical committees approved protocol we collected data from 17 patients seen at the emergency department of Froedtert Hospital, Milwaukee, Wisconsin. The sample set had a distribution of hemoglobin levels in three different groups: ¡9 g/dl (N=3), 9-13 g/dl (N=6), 13-16 g/dl (N=8). We found significant concordance of the red pixel intensity by mini-video and the hemoglobin level by gold standard venipuncture laboratory testing with a linear least squared regression correlation coefficient (r) of 0.68. We know that red pixel intensity is influenced by level of oxygenation which in all of our patients was high. In exploring our data we surmised that red pixel color intensity at the fingertip might be influenced by the thickness of the skin which in turn is age and gender-related (men have thicker skin, and thickness increases with age).

9.2.6 Evaluation: Evaluation of Machine Learning Algorithms For Application in Clinical Setting

This is study was performed for developing insights on machine learning algorithm development techniques for multi-parameter patient monitoring for the goal of evaluating these algorithms. Traditional devices that monitor patients vital signs such as heart rate, blood pressure, oxygen saturation or respiratory rate in the hospital settings have their own individual warning systems for clinical intervention. But these devices do not consider the context of the patients status and the relative values of other vital signs. It is argued in medical science and machine learning community that data fusion techniques including machine learning algorithms that consider multiple vital signs would improve early warning systems by reducing number of false alarms and at the same time would represent a better understanding of the patients underlying physiological status. Questions remain how to evaluate these algorithms such that, the system can be used in a clinical setting; such as

- How to select the training, testing and validation data set for developing the algorithms?
- Do the algorithms outperform the existing reference system for early warning score?
- And, if the algorithms have less number of false alarms and thus help solve the problem of alarm fatigue.

9.3 Broader Impact

The work would help understand the different issues related to building sensor based systems and developing algorithms for interpreting clinical events. These systems would also help the under served groups for the improvement of quality of life. Thus would vastly increase the use of existing infrastructure for low cost solutions of health care problems with the goal of improved patient outcome.

BIBLIOGRAPHY

[1] https://itunes.apple.com/us/app/digifit-icardio-multi-sport/id314841648?mt=8.

[2] https://play.google.com/store/apps/details?id=si.modula.android.instantheartrate&hl=en.

[3] http://www.personalizedmedicinecoalition.org/about.

[4] http://www.personalizedmedicinecoalition.org/about/about-personalized-medicine/the-case-for-personalized-medicine.

[5] National early warning score (news): Standardising the assessment of acute-illness severity in the nhs. In *Report of a Working Party*, 2012.

[6] Scully, C., Lee, J., Meyer, J., Gorbach, A., Granquist-Fraser, D., Mendelson, Y., Chon, K. Physiological paramter monitoring from optical recordings with a mobile phone. In *IEEE Transactions on Biomedical Engineering*, volume 59, 2012.

[7] Ashraf, A., Lucey, S., Cohn, J., Chen, T., Ambadar, Z., Prkachin, K.,Solomon, P., Eheobald, B. . The painful face - pain expressino recognition using active appearancs models. In *ICMI*, 2007.

[8] Ashraf, A., Lucey, S., Cohn, J., Chen, T., Prkachin, K., Solomon, P. The painful face ii– pain expression recognition using active appearance model. In *International Journal of Image and Vision Computing*, volume 12, pages 1788–1796, 2009.

[9] Au, E., Loprinzi, C., Dhodapkar, M., Nelson, T., Novotny, P., Hammack, J., O'Fallon, J. Regular use of verbal pain scale improves the understanding of oncology inpatient pain intensity. In *J Clin Oncol*, volume 12, pages 2751–2755, 1994.

[10] Ayzenberg, Y., Hernandez, J., Picard, R. Feel: Frequent eda and event logging - a mobile social interaction stress monitoring system. In *CHI*, 2012.

[11] Baruch, M., Warburton, D., Bredin, S., Cote, A., Gerdt, D., Adkins, C. Pulse decomposition analysis of the digital arterial pulse during hemorrhage simulation. In *Nonlinear Biomed. Phys.*, volume 5, 2011.

[12] Becouze, P., Hann, C., Chase, J., Shaw, G. Measuring facial grimacing for quantifying patient agitation in critical care. In *Computer Methods and Programs in Biomedicine*, volume 87, pages 198–147, 2007.

[13] Bellomo, R., Goldsmith, D., Uchino, S., Buckmaster, J., Hart, G., Opdam, H., Silvester, W., Doolan, L., Gutteridge, G. A prospective before-and-after trial of a medical emergency team. In *MJA Rapid Online Publication*, 2003.

[14] Berkow, L., Rotolo, S., Mirski, E. Continuous noninvasive hemoglobine monitoring during complex spine surgery. In *Anesth Analg*, volume 113, pages 1396–1402, 2011.

[15] Boehner, K., DePaula, R., Dourish, P., Sengers, P. How emotion is made and meassured. In *International Journal of Human Computer Studies*, volume 65, 2007.

[16] Boulos, M. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. In *Biomedical Engineering Online*, 2011.

[17] Bradley, M., Miccoli, L., Escrig, M., Lang, P. The pupil as a measure of emotional arousal and autonomic activation. Center for the Study of Emotion and Attention, University of Florida, 2008.

[18] Brahman, S., Chuang, C., Shih, F., Slack, M. Machine recognition and representation of neonatal facial displays of acute pain. In *Artif. Intel. Med.*, volume 36, pages 211–222, 2006.

[19] Brahnam, S., Nanni, L., Sexton, R. Introduction to neonatal facial pain detection using common and advanced face classificaton techniques. In *Stud. Comput. Intel*, volume 48, pages 225–253.

[20] Bruera, E., Kuehn, N., Miller, M., Selmser, P., Macmillan, K. The edmonton symptom assessment system (esas): a simple method of the assessment of palliative care patients. In *J Palliative Care*, volume 7, pages 6–9, 1991.

[21] Cannon, C., Braxton, C., Kling-Smith, M., Carlton, E., Moncure, M. Utility of the shock index in predicting mortality in traumatically injured patients. In *J. Trauma*, volume 67, pages 1426–1430, 2009.

[22] Causey, M., Miller, S., Foster, A., Beekley, A., Zenger, D., Martin, M. Validaton of noninvasive hemoglobin measurements using the masimo radical -7, sphb station. In *Am J Surg*, volume 201, pages 592–8, 2011.

[23] Chon, K., Dash, S., Kihwan, J. Estimation of respiratory rate from photoplethysmogram data using time frequency spectral estimation. In *IEEE Transactions on Biomedical Engineering*, volume 56, 2009.

[24] Cleeland, C., Ryan, K. Pain assesment: global use of the Brief Pain Inventory. In *Annals Academy Med Singapore*, volume 23, pages 129–138, 1994.

[25] C. D. P. M. W. P. Clifton, L. and L. Tarassenko. Gaussian processes for personalized e-health monitoring with wearable sensors. In *IEEE Transaction on Biomedical Engineering*, volume 60, 2013.

[26] Cooke, W., Rickards, C., Ryan, K., Convertino, V. Autonomic compensation to stimulated hemorrhage with heart period variability. In *Critical Care Medicine*, volume 36, 2008.

[27] Craig, K., Grunau, R., Aquan-Assee, J. Judgment of pain in newborns: facial activity and cry as determinants. In *Can J Behave Sci*.

[28] Craig, K., Hadjistavropoulos, H., Grunau, R., Whitefield, M. A comparison of two measures of facial activity during pain in the newborn child. In *J Pediatr Psychol*, volume 19, pages 305–18, 1994.

[29] Craig, K, Whitfield, R., Grunau, R., Linton, J., Hadjistavropoulos, H. Pain in the preterm neonate: behavioral and physilogical indices. In *Pain*, volume 52, pages 287–299, 1993.

[30] Ekman,P., Friesen, W. . Consulting Psychologist Press, 1978.

[31] Fjeldsoe, B., Alison, B., Marshall, L., Miller, Y. Behaviour Change Interventions Delivered by Mobile Telephone Short-Message Service. In *American Journal of Preventing Medicine*, 2009.

[32] Frasca, D., Dahyot-Fizelier, C., Catherine, K., Levrat, Q., Debaene, B., Mimoz, O. Accuracy of a continuous hemoglobin monitor in intensive care unit patients. In *Crit Care Med*, volume 39, pages 2277–82, 2011.

[33] Fujiki, Y. iPhone as a physical activity measurement platform. In *CHI 2010(Student Research Competition)*, 2010.

[34] Gayat, E., Aulagnier, J., Matthieu, E., Boisson, M., Fishcler, M. Noni-invasive measurment of hemoglobin: assessment of two different point-or-care technlogies. In *PLoSone*, volume 7, 2012.

[35] Gayat, E., Bodin, A., Sportiello, C., Boisson, M., Dreyfus, J., Mathieu, E., Fishcler, M. Performance evaluation of a noninvasive hemoglobin monitoring device. In *Ann Emerg Med*, 2011.

[36] Goldberger, A., Amaral, L., Glass, L., Jausdorff, J., Ivanov, P., Mark, R., Mietus, E., Moody, G., Peng, C., Stanley, H. In *Circulation*, volume 101, pages 215–220, 2000.

[37] Grunau, R., Craig, K. Pain expression in neonates: facial action and cry. In *Pain*, volume 28, pages 395–410.

[38] Grunau, R., Johnston, C., Craig, K. Neonatal facial and cry responses to invasive and non-invasive procedures. volume 42, pages 295–305, 1990.

[39] Gutierrez, G., Reines, H., Wulf-Gutierrez, M. Clinical review: hemorrhagic shock. In *Critical Care*.

[40] Hadar, E., Raban, O.,Bouganim, T. . Precision and accuracy of non invasive hemoglobin measurements during pregnancy. In *Helen Schneider Hospital for Women, Rabin Medical Center, Petach Tikva and the Sackler Faculty of Medicine, Tel Aviv University*.

[41] Hadjistavropoulos, H., Craig, K. Acute and chronic low back pain: Cognitive, affective, and beahvioral dimensions. In *Journal of Consulting and Clinical Psychology*, volume 62, page 341, 1994.

[42] Hansen, F., Christensen, S. R. Emotions, Advertising and Consumer Choice. Copenhagen Business School Press, 2007.

[43] Haque, M., Kawsar, F., Adibuzzaman, M., Ahamed, S., Love, R., Dowla, R., Roe, D., Hossain, S., Selim, R. Findings of e-ESAS: a mobile based symptom monitoring system for breast cancer patients in rural Bangladesh. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, volume 293, May 2012.

[44] Healey, J. Recording affect in the field: Towards methods and metrics for improving ground truth labels. In *Proc. of the Affective Computing and Intelligent Interactions*, 2011.

[45] Healey, J., Nachman, L., Subramanian, S., Shahabdeen, J., Morris, M. Out of the Lab and into the Fray: Towards modeling emotion in everyday life. In *Proc. 8th International Conference on Pervasive Computing*, 2010.

[46] Hravnak, M., Edwards, L., Clontz, A., Valenta, C., DeVita, M., Pinsky, M. Defining the incidence of cardiorespiratory instability in patients in step-down units using an electronic integrated monitoring system. In *Arch Intern Med*, volume 168, 2008.

[47] Isbister, K. How to stop beign a buzzkill: designing yamove! a mobile tech mash-up to truly augment social play. In *MobileHCI*, 2012.

[48] James, A., Sebe, N. Multimodal human computer interaction: A survey. In *Computer Vision and Image Understanding*, volume 108, October 2007.

[49] Joliffe, I. Principal Component Analysis. In *Springer*, 2002.

[50] Jonathan, E., Leahy, M. Investigating a smartphone imaging unit for photoplethysmography. In *Physiological measurements*, volume 31, 2010.

[51] Jonathan, E., Leahy, M. Cellular phone based photoplethysmographic imaging. In *Biophotonics*, 2011.

[52] Kaplan, W. Can the ubiquitous power of mobile phones be used to improve health outcomes in developing countries? In *Globalization and Health*, 2006.

[53] Kapoor, A., Picard, R. Multimodal Affect recognition in learning environments. pages 677–682, 2005.

[54] Kauvar, D., Lefering, R., Wade, C. Impact of hemorrhage on trauma outcome: an overview of epidemeiology, clinical presentations, and therapeutic considerations. In *J. Trauma*, volume 141, 2006.

[55] Kording, K., Wolpert, D. Bayesian decision theory in sensorimotor control. In *Trends in Cognitive Sciences*, pages 319–326, 2006.

[56] Kramer, C., Kinsky, M., Prough, D. Salinas, J. . Closed-Loop Control of Fluid Therapy for Treatment of Hypovolemia. In *J Trauma*, volume 64, 2008.

[57] Lamhaut, L., Apriotesei, R., Combes, X., Lejay, M., Carli, P., Vivien, B. Comparison of the accuracy of noninvasive hemoglobin monitoring by spectrophotometry (sphb) and hemocue with automated laboratory hemoglobin measurement. In *Anesthesiology*, volume 3, page 115, 2011.

[58] Littlewort, G., Bartlett, M., Lee, K. Faces of pain: Automated measurement of spontaneous facial expressions of genuine and posed pain. In *13th Joint Symposium on Neural Computation, San Diego, CA*, 2006.

[59] Littlewort, G., Bartlett, M., Lee, K. Automatic coding of facial expression displayed during posed and genuine pain. In *Image and Vision Computing*, volume 12, pages 1741–1844, 2009.

[60] Lucey, P., Cohn, J., Prkachin, K., Solomon, P, Chen, S. . Painful monitoring: Automatic pain monitoring using the unbc-mcmaster pain expression archive database. In *Image and Vision Computing*, pages 197–205, 2012.

[61] Macknet, M., Allard, M., Applegate, R., Rook, J. The accuracy of noninvasive and continuous total hemoglobin measurement by pulse co-oximetry in human subjects undergoing hemodilution. In *Anesth Analg*, volume 111, pages 1424–1426, 2010.

[62] Merrill, S. Markov chain methods in the analysis of heart rate variability. In *Fields Institute Communications*, volume 11, 1997.

[63] Merrill, S. J. Markov chains for identifying nonlinear dynamics. In *Nonlinear Dynamical Systems Analysis for the Behavioral Sciences Using Real Data*, 2011.

[64] Metallinou, A., Narayanan, S., Lee, S. Decision level combination of multiple modalities for recognition and analysis of emotional expression. In *Proc. of the International Conference on Acoustics, Speech, and Signal Processing*, 2010.

[65] Miller, R., Ward, T., Shiboski, S., Cohen, N. A comparison of three methods of hemoglobin monitoring in patients undergoing spine surgery. In *Anesth Analg*, volume 112, pages 858–863, 2011.

[66] Monwar, M., Prkachin, K., Rezaei, S. Eigenimage based pain expression recognition. In *International Journal of Applied Mathematics*, May 2007.

[67] Monwar, M., Rezaei, S. Pain recognition using artificial neural network. In *IEEE International Symposiym on Signal Processing and Information Technology*, 2006.

[68] Monwar, M., Rezaei, S., Prkachin, K. Appearance based pain recognition from video sequences. In *IAENG International Journal of Applied Mathematics*, pages 2429–34, 2006.

[69] Nguyen, B., Vincent, J., Nowak, E., Coat, M., Paleiron, N., Gouny, P., Ould-Ahmed, M., Guillouet, M., Arvieux, C., Gueret, G. . The accuracy of noninvasive hemoglobin measurement by multiwavelength pulse oximetry after cardiac surgery. In *Anesth Analg*, 2001.

[70] Nicolaou, A., Pantic, M., Gunes, H. Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. In *IEEE Transactions On Affective Computing*, volume 2, 2011.

[71] Niese, R., Al-Hamadi, A., Panning, A., Brammen, D., Ebmeyer, U., Michaelis, B. Towards pain recognition in post-operative pahses using 3d-based features from videos and support vector machines. In *International Journal of Digital Content Technology and its Applications*, volume 3, pages 21–31, 2009.

[72] Pantic, M., Roisman, G., Huang, T., Zeng, Z. A survey of affect recognition methods: Audio, visual and spontaneous sxpressions. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 31, January 2009.

[73] Picard, R. Affective computing: Challenges. In *International Journal of Human-Computer Studies*, 2003.

[74] Picard, R. W. In *Affective Computing*. The MIT Press, 1997.

[75] Poh, M.-Z., McDuff, D. J. ,Picard, R. W. . Advancements in Noncontact. Multiparameter Physiological Measurements Using a Webcam. In *IEEE Transactions on Biomedical Engineering*, volume 58, 2011.

[76] Prkachin, K. The consisteny of facial expressions of pain: a comparison across modalities. In *Pain*, pages 297–306, 1992.

[77] Prkachin, K. Assessing pain by facial expression: facial expression as nexus. In *Pain Research and Management: The Journal of the Canadian Pain Society*, volume 14, pages 53–58, 2009.

[78] Prkachin, K., Craig, K. Influencing non-verbal expressions of pain: Signal detection analyses. In *Pain*, volume 21, pages 399–409, 1985.

[79] Prkachin, K., Mercer, S. Pain expression in patients with shoulder pathology: validity, coding properties and relation to sickness impact. volume 39, pages 257–65, 1989.

[80] Prkachin, K., Rezaei, S. The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. In *Pain*, volume 13, pages 267–274, 2009.

[81] Rabbi, M., Ali, S., Choudhury, T., Berke, E. Passive and in-situ assessment of mental and physical well-being using mobile sensors. In *Proc. of the 13th International Conference on Ubiquitous computing*, 2011.

[82] Rady, M., Rivers, E., Nowak, R. Resuscitation of the critically ill inthe ed: responses of blood pressure, heart rate, shock index, central venous oxygen saturation, and lactate. In *Am. J. Emerg. Med.*, volume 14, pages 218–225, 2009.

[83] Raheja, J., Kumar, U. Human facial expression detection from detected in capture image using back propagation neural network. In *Internationl Journal of Computer Sceince and Information Technology*, pages 116–123, 2010.

[84] Rime, B., Mesquita, B., Philipot, P. Long lasting cognitive and social consequences of emotion: Social sharing and rumination. In *European Review of Social Psychology*, 1992.

[85] Rime, B., Philippot, P., Zech, E., Luminet, O., Finkenauer, C. Social sharing of emotion: New evidence and new questions. In *European review of social psychology*, volume 9, pages 145–189.

[86] Russell, J. A circumplex model of affect. In *Journal of Personality and Social Psychology*, volume 39, 1980.

[87] Ryan, K., Rickards, C., Ludwig, D., Convertino, V. Tracking central hypovolemia with ecg in humans: cautions for the use of heart period variability in patient monitoring. In *Shock*, volume 33, pages 583–589, 2010.

[88] S. Grossman, S.,Nesbit, S. In *Cancer Pain in Clinical Oncology, 3rd Edition*. Elsevier, Philadelphia, 2004.

[89] Sandby-Moller, J., Pulsen, T., Wulf, H. Epidermal thickness at different body sites: relationship to age, gender, pigmentation, blood content, skin type and smoking habits. In *Acta Derm Venereol*, volume 83, pages 410–3, 2003.

[90] Scherer, K., Ekman, P. Methods for measuring facial action. In *Handbook of Methods in Nonverval Behavior Research*, pages 45–135. Cambridge University Press, 1982.

[91] Serre, T., Bouvrie, J., Ivanov, Y. Error weighted classifier combination for multimodal human identification. In *Tech. Rep., MIT, Cambridge, MA*, 2005.

[92] Sharma, R., Huang, T., Pavlovic, V. Toward multimodal human computer interface. In *Proc. IEEE*, volume 86, pages 853–869, 1998.

[93] Tarassenko, L., Hann, A., Young, D. Integrated monitoring and analysis for early warning or patient deterioration. In *British Journal of Anaesthesia*, volume 97, 2006.

[94] Thayer, J., Yamamoto, S., Brosschot, J. In *International Journal of Cardiology*, volume 141, pages 122–131.

[95] Tian, Y., Cohn, J., Kanade, T. Facial Expression Analysis. In *Handbook of Face Recognition*. Springer, 2005.

[96] Tian, Y., Cohn, J., Kanade, T. Affect detection: An interdisciplinary review of models, methods, and their applications. In *IEEE Transactions on Affective Computing*, volume 1, 2010.

[97] Turk, D., Melzack, R. In *Handbook of Pain Assessment, New York*, 2001.

[98] Turk, M., Pentland, A. Eigenfaces for recognition. In *Journal of Cognitive Neuroscience*, volume 3, 1991.

[99] Twycross, R., Fairfield, S. Pain in far-advanced cancer. In *Pain*, volume 14, pages 303–310, 1982.

[100] Verkruysse, W., Svaasand, L., Nelson, J. Remote plethysmographic imaging using ambient light. In *Opt Express*, 2008.

[101] Wagner, J., Andre, E., Kim, N.J. From physiological signals to emotions implementing and comparing selected methods for feature extraction and classification. In *IEEE Int'l Conf. Multimedia and Expo*, pages 940–943, 2005.

[102] Weiser, M. Some computer science issues in ubiquitous computing. In *Communications of the ACM*, pages 74–83, 1993.

[103] Yambor, W., Draper, B., Beveridge, J. Analyzing pca based face recognition algorithms: Eigenvector selection and distance measures. In *Second Workshop on Empirical Evaluation Methods in Computer Visio*, 2000.

[104] Zeng, Z., Pantic, M., Roisman, G., Huang, T. A survey of affect recognition methods: audio, visual and sponetaneous expressions. In *IEEE Transactions of pattern analysis and machine intelligence*, volume 31, 2009.

[105] Zhu, X. Face detection, pose estimation and landmark localization in the wild. 2012.

[106] Zong, W., Heldt, T., Moody, G., Mark, R. An open source algorithm to detect onset of arterial blood pressure pulses. In *Computers in Cardiology*, pages 259–262, 2003.