COMPUTER VISION APPROACH FOR ESTIMATING HUMAN HEALTH

PARAMETERS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Mayank Gupta

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science in Industrial Engineering

December 2018

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. Vaneet Aggarwal

     School of Industrial Engineering

Dr. Denny Yu

     School of Industrial Engineering

Dr. Roshanak Nateghi

     School of Industrial Engineering

Approved by:

     Dr. Steven Landry

          Head of the School Graduate Program

Dedicated to my parents - Mrs. Veena Gupta and Mr. Shyam Sunder Gupta.

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my major advisor, Dr. Vaneet Aggarwal, for his excellent guidance and support during my MS study. He has always been very helpful, accommodating and understanding.

I would also like to thank Dr. Denny Yu, for his guidance throughout the course of my stay at Purdue and, support whenever I needed it. I also would like to thank Dr. Roshanak Nateghi for providing suggestions, help and resource materials whenever I needed them.

Finally, I would like to thank family members for their support throughout this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Gupta, Mayank MSIE, Purdue University, December 2018. Computer Vision Approach for Estimating Human Health Parameters. Major Professor: Vaneet Aggarwal.

Measurement of vital cardiovascular health attributes, e.g., pulse rate variability, and estimation of exertion level of a person can help in diagnosing potential cardiovascular diseases, musculoskeletal injuries and thus monitoring an individual's well-being. Cumulative exposure to repetitive and forceful activities may lead to musculoskeletal injuries which not only reduce workers' efficiency and productivity, but also affect their quality of life. Existing techniques for such measurements pose a great challenge as they are either intrusive, interfere with human-machine interface, and/or subjective in the nature, thus are not scalable. Non-contact methods to measure these metrics can eliminate the need for specialized piece of equipment and manual measurements. Non-contact methods can have additional advantages since they are potentially scalable, portable, can be used for continuous measurements, and can be used on patients and workers with varying levels of dexterity and independence, from people with physical impairments, shop-floor workers to infants. In this work, we use face videos and the photoplethysmography (PPG) signals to extract relevant features and build a regression model that can predict pulse rate, and pulse rate variability, and a classification model that can predict force exertion levels of 0%, 50%, and 100% (representing rest, moderate effort, and high effort), thus providing a non-intrusive and scalable approach. Efficient feature extraction has resulted in high accuracy for both the models.

1. INTRODUCTION

Regular and non-invasive measurement of important health attributes such as pulse rate (PR), pulse rate variability (PRV), and blood pressure (BP), force exertion level are important due to their fundamental role in tracking one's fitness level, diagnosis of cardiovascular diseases, and monitoring of well-being. Such measurements are essential to obtain warning signs for heart diseases, anxiety, fatigue, and musculoskeletal disorders (MSDs) at home, while working in the office, or during heavy-lifting jobs such as in manufacturing shop floors. It becomes really imperative if these frequent measurements can be done passively using non-contact methodologies. This work explores the use of facial features from the videos of human to predict these vital health attributes.

This work proposes the use of computer vision and machine learning technique that can predict health related metrics such pulse rate (PR), pulse rate variability (PRV) and force exertion level of the person using the features extracted from the face video. Pulse rate is the number of times per unit time that the arteries expand and contract in response to the heart, and is usually expressed in beats per minute (bpm). Several studies [1–4] suggest that the heart rate of a normal adult is in the range of 50-90 bpm while during sleep heartbeat of around 40-50 bpm is considered normal. However, abnormalities in the heart is considered as a key factor for heart disease and heart failure [5]. Pulse rate variability is the variation in the time interval between two expansions of the artery. It is usually measured by the variation in beat-to-beat interval. This metric is considered as a non-invasive technique for measuring autonomic nervous system (ANS) activity [6]. The autonomic nervous system has two branches; sympathetic nervous system (SNS) and parasympathetic nervous system (PNS) and is regulated by hypothalamus. Its function includes control of respiration, cardiac regulation, vasometer activity and certain reflex actions like coughing, sneez-

ing, swallowing, and vomiting. High-frequency (HF) component of PRV is affected by efferent vagal (parasympathetic) activity and it decreases during the conditions of acute time pressure, emotional strain, mental stress, and elevated anxiety [7–9]. The low-frequency (LF) component of PRV is known to contain both sympathetic and vagal influences [10]. In the United States, 155 million people work full-time as part of their daily lives. Although all employers are ethically required to provide a safe and healthy workplace for their employees, people are still getting hurt daily. Workplaces injuries like musculoskeletal disorders are preventable, and workplace risk factors are known. However, monitoring these factors reliably and in a scalable way is a key challenge. As per [11], MSDs such as strains and sprains that occur due to overexertion constitutes of 349,050 cases that poses a huge burden on American workers and industries. Overexertion has been known to be the most promising cause of disabling injury as reported by the 2017 Liberty Mutual Workplace Safety Index report that causes a huge direct cost of around $ 13 billion. There are multiple factors that contributes towards MSDs, but physical work demands has always been an important contributor for MSDs. Furthermore, they not only impact individual worker's health and quality of life [12], they also result in significant cost to employees and society (e.g., workers compensation, medical care, loss productivity, training temporary workers). [13–19]. Therefore, it leads to the necessity of measuring PR, PRV, and exertion level frequently and accurately as it provides critical signs of one's well being and any abnormality could lead to potential health problems.

High force exertion levels are reported as the most common contributing factors with sufficient evidence to suggest a causal relationship for work-related musculoskeletal disorders (MSDs) [20–24]. A comprehensive report by the National Institute for Occupational Safety and Health (NIOSH) lists high/sustained force, repetitive movements, and poor biomechanical postures are contributors to MSDs, with conclusion that evidence exists linking force to musculoskeletal injuries [25].

Several key physiological and biomechanical mechanisms are proposed for how force exertions lead to injuries. For instance, chronic low back pain can be a result

of tears in the soft tissues [26]. For instance, high and/or frequent force exertions initiates lumbar disc damage and degeneration [27]. In addition to high force exertions, prolonged/sustained force exertions could also lead to work-related MSDs. For example, prolonged force exertions could lead to wrist injuries where frequent force exertions by the hand (e.g., pinching and griping) lead to and exacerbate inflammation of the carpal tunnel cumulative tissue stress can eventually lead to injuries [28].

Currently, the gold standard techniques for measuring PR, and PRV include using intrusive contact devices such as electrocardiogram (ECG), chest straps, and pulse oximeters. Traditionally, ECG was extensively used for such measurement but recent trend has been shifted towards using pulse oximeters because of its low cost. Pulse oximeters depends on photoplethysmography (PPG) signal for the measurement of PR and PRV. These devices are generally attached to the skin surface, usually on fingertip or earlobe for such measurements. PPG is a low cost, non-invasive optical technique for volumetric measurement of an organ. It utilizes opto-electronic components: a source of light to illuminate the skin and a photodetector to measure the small variation in the light intensity. PPG detects the blood volume change inside the arteries when the skin is illuminated by the light source. The light when projected on the skin surface travels through the skin, arteries, and blood vessels. There are two type of reflected lights received by the pulse oximeter: surface reflections and sub-surface reflections. The reflection from the skin is surface reflection and is constant whereas the reflection from blood inside the vessels changes because of the change in blood volume with each cardiac cycle. The intensity of the sub-surface reflected light carries the information that can be utilized to calculate the required health parameters. The change in the volume of blood in arteries and capillaries is in synchronization with the cardiac cycle. PPG waveform comprises of two components, namely the AC and the DC component. AC component comprises of a pulsatile physiological waveform caused by the changes in the blood volume with each heart beat and DC component is slow varying and contains low frequency components associated with respiration, sympathetic nervous system, and thermoregulation [29].

Therefore, extraction of PPG waveform is very important to measure the well being of the individual. Although pulse oximeter is easy to use, they have limitations for frequent measurements. First, it requires the purchase of equipment and needs either the health provider or the user to manually perform the measurements. Second, the device needs to be carried to the different places that the user goes, limiting its use. Third, the finger clip-on and earlobe clip-on may not always fit well on every individual due to varying size of fingers and earlobes. The improper fitting of the device may lead to estimation errors [30]. Fourth, using clip on may be potentially uncomfortable during long use.

Although repetition, postures, and vibration are contributors to injuries, **force** is one of the hardest to measure because it is difficult to observe and depend on individual's effort. For example, changes in expressions are subtle unless high forces and strong efforts are needed. Many methods are currently available to measure the force exertion levels. However, each method vary in reliability and feasibility as they are either 1) intrusive (e.g., disrupts the worker while they are performing their job), 2) interfere with human machine interface (e.g., need to install force gauges on tool-handles and machine controls), 3) subjective, and most importantly 4) not widely scalable across all workers, jobs, and workplaces as trained ergonomics and safety professionals are needed to implement these methods.

Various methods have been used by different researchers to measure, estimate workers hand force exertions. For example, the physical exertion level is commonly rated using visual scales [31]. Other techniques include estimating hand forces through context, i.e., using the object weights or checking the carrying loads by observing and interviewing the workers [32], measured with a force gauge or mimicked on a hand dynamometer by workers [33, 34], observed by ergonomists, or measured by electromyography on the forearm muscles [35, 36].

The conventional method to measure hand grip force uses hand grip dynamometers [33, 34]. Most hand dynamometers operate using strain gauge and directly measures hand grip strength. In workplaces, the workers may be asked to replicate the

tasks forces on grip dynamometers. Although these devices provide actual force measurements, their usability is limited due to their availability to workers.

In observational methods, the force levels are observed and estimated by trained ergonomists. The ergonomists are trained to recognize these subtle cues (twist in body, strain in face, perspiration). These signs will become more clear in high force assessing requirements. This method could be also performed on recorded videos of the workers. This method is subjective and based on the estimations [37].

Electromyogram (EMG) is a signal that can be measured from the skin surface. Various studies have used the EMG sensors to measure the muscle activation and hand grip strength. The EMG signals measures the activation of forearm muscles. The recorded signals can be filtered and normalized with the maximum activity to represent the hands grip forces [35, 36, 38]. This method requires the EMG sensors which are not widely available, cannot be used in workplaces due to time constrains, and are intrusive to workers.

Thus, this work considers the use of non-contact based robust approach where passive video of the face can be used to estimate the health metrics of the individual without the need of any specialized piece of equipment.

Monitoring of health parameters like pulse rate and pulse rate variability using non-contact methods like videos from camera has been recently considered in [39–46], where it is shown that the PPG signal can be extracted from the videos of the face.It can be done without any dedicated source of light with the help of low cost digital camera. The non-contact measurement using camera's video could find many applications including determination of health parameters of people working in an office environment, shop-floor, newborn infants in the hospital where using contact probes may not be possible. These applications are promising because of the availability of cameras in these places. In these works, the PPG signal is extracted from each individual, and thus the coefficients of the video features that provide the PPG signal are dependent on the individual. In contrast, we do not consider individual characteristics in the prediction. The proposed method can thus help predict health metrics

of an individual for which no training sample has been collected in the past, making our methodology robust. Also, the use of non-contact methodology has an additional advantage of being scalable, and portable since the presence of camera is ubiquitous in all electronic gadgets carried by the users.

The novel approach proposed in this work has three key steps. The first step considers designing the experiments to gather the data. This involves capturing the video of the subjects under different conditions. Second step involves extracting the "right" features from different frames of the video. Once the features corresponding to the face in each frame are obtained, the third step includes training a deep neural network to learn health parameters from the above obtained features.

1.1 Key Contributions

The key contributions in predicting cardiovascular parameters and force exertion level are summarized here:

1.1.1 Cardiovascular Parameters:

- One of the key contribution is designing an experiment in order to collect the data that can help in predicting pulse rate and pulse rate variability from the facial videos. To the best of our knowledge, there is no existing data-set that contains such information required to train the machine learning model.

- Most of the work that has been done in providing non-contact mechanism for estimating pulse rate and pulse rate variability doesn't involve using machine learning approach. Most of the methods are dependent on using techniques of signal processing, and other domain knowledge to estimate cardiovascular parameters passively. There have been certain work that uses machine learning, but none of the work predicts pulse rate variability using machine learning. This work uses the data collected from the designed experiment and extends the use

of machine learning in such prediction. The approach has been evaluated on a group of people with different ethnicities, where we find that the proposed approach has lower than 7.4% error in predicting each of the pulse rate, low- and high- frequency components of the pulse rate variability. The proposed approach also reduces the root mean squared error in predicting pulse rate variability by 83.3% as compared to the state-of-the-art.

1.1.2   Force Exertion Level

- This work provides first ever approach for predicting the force exertion level using techniques of computer vision and machine learning, to the best of our knowledge. The proposed algorithm classifies between three different levels of force exertion, i.e., 0%, 50%, & 100%. The methodology provides an overall accuracy of over 80% in correctly classifying these three levels. This algorithm uses efficient feature extraction methods from the facial videos and the PPG signals to perform the classification. The proposed method first uses a classification based on features extracted from the video data to classify between two levels, 100% and $\leq 50\%$. The features from the PPG signals are then used to differentiate between 0% and 50%. Since the PPG signals can be obtained from the face videos [42], this approach is a computer vision approach with the face video as an input and an estimate of force exertion level as the output. Since we are not aware of an existing data set containing face videos and the force exertion levels, we design our own experiment to collect the relevant data for training our model. The data is collected using 20 subjects in total, where each subject was asked to perform different levels of force exertion activities and their videos & PPG data was recorded. Relevant set of features are extracted from this data and two neural networks are trained to achieve an overall accuracy of 81.7%.

- The second key contribution in our work is the extraction of average movement of facial landmarks from the video data. Such features has never been used to train a machine learning algorithm for predicting force level in the existing literature, to the best of our knowledge. The feature extraction is essential since the large number of features in the face video may cause the classification algorithm to overfit to training samples and generalize poorly to new samples. It has been demonstrated that the extracted features are the key features in training the model to provide the classification between high (100%) & low (0% & 50%) force exertion level. Deepface algorithm [47] has been used to process the collected videos. The different frames of the videos are aligned and human face is detected in each frame using Deepface. The spatial location of facial landmarks (128 points on the face) are tracked in the entire video and the average movement of each landmark is calculated relative to its position in the first frame of the video. The detailed explanation on such feature extraction is given in Section 4.2.2.

- The third contribution in the proposed paper includes efficient feature extraction from photoplethysmogram (PPG) signals. PPG signal is collected for each subject at every force exertion level during our experiments. PPG signal has been utilized to extract various kind of features as detailed in Section 4.2.2. This approach of feature extraction is novel, and has not been studied in prior works to the best of our knowledge. The extracted PPG features provide second level of classification between 0% and 50% force exertion levels in our work. For the cases where the face video signals predict the low levels (0% or 50%), the extracted features from the PPG signals are used to obtain efficient classification.

- The proposed approach is shown to be robust to unseen data from an activity level that is different from the activities in our experiment. This is done by recording the face video and the PPG signals of the subjects at a new activity

level corresponding to when the subject is talking. The average movement landmarks for talking were extracted from the videos, and were used to predict force exertion level from our first model. If the first model predicts low, the extracted PPG features from the obtained PPG signal are passed through our second trained model. The results shows that first model classifies 7 out of 7 subjects belonging to low (0% & 50%) force exertion level category and second model predicts 0% force exertion level for 5 out of 7 subjects

The rest of the chapters are organized as follows: The discussion on the related work is presented in chapter 2. The experiments conducted to gather the data is explained in chapter 3. The novel approach adapted to process videos, extract features, and train neural networks is explained in 4. The results obtained in this work are demonstrated in chapter 5. Chapter 6 provides additional discussions for the obtained results and Chapter 7 concludes the thesis with a brief mention of the potential future work.

2. LITERATURE STUDY

Over the past decade, there has been interest in developing methodologies for measuring health metrics like pulse rate, pulse rate variability, blood pressure, and blood oxygen level using non-contact devices (e.g., camera videos) as opposed to contact devices [39–42, 48–51]. Although pulse oximeter device provides easier method for measuring PR and PRV than using ECG, but non contact method has its own benefits as discussed in chapter 1.

The authors in [49] showed that it is possible to extract PPG signal from the video using complementary metal-oxide semiconductor camera by illuminating a region of tissue using through external light emitting diodes at dual wavelength (760nm and 880nm). Further, the authors of [39] demonstrated that PPG signal can be estimated by just using ambient light as a source of illumination along with simple digital camera. Further in [40], PPG waveform was estimated from the videos recorded using low cost webcam. The red, green, and blue channels of the images were decomposed into independent sources using independent component analysis. One of the independent source was selected to estimate PPG and further calculate PR, and PRV. All these works showed the possibility of extracting PPG signal from the videos and proved the similarity of this signal with the one obtained using contact device.

The authors of [42] proposed a methodology that overcomes a challenge in extracting PPG for people with dark skin tones. The challenge due to slight movement and low lighting conditions during recording a video is also addressed. They implemented the method where PPG signal is extracted from different regions of the face and signal from each region is combined using their weighted average making weights different for different people depending on their skin color.

There are other attempts where authors of [43–45, 52] have introduced different methodologies to make algorithms for estimating pulse rate robust to illumination

variation and motion of subjects. The paper [45] introduces a chrominance-based method to reduce the effect of motion in estimating pulse rate. The authors of [44] uses a technique in which face tracking and Normalized Least Square adaptive filtering is used to counter the effects of variations due to illumination and subject movement. The paper [52] resolves the issue of subject movement by choosing the rectangular ROI's on the face relative to the facial landmarks and facial landmarks are tracked in the video using pose-free facial landmark fitting tracker discussed in [46] followed by the removal of noise due to illumination to extract noise-free PPG signal for estimating pulse rate.

Recently, use of machine learning in prediction of health parameteres have gained attention. The paper [53] used supervised learning methodology to predict the pulse rate from the videos taken from any off-shelf camera. The authors trained the support vector machine with radial basis function. Their model showed the possibility of using machine learning methods to estimate pulse rate. However, our method outperforms their results when root mean squared error of the predicted pulse rate is compared which is discussed later in the paper. The authors in [54] proposed a deep learning methodology to predict pulse rate from the facial videos. The researchers trained a convolutional neural network (CNN) on the images generated using Short-Time Fourier Transform (STFT) applied on the R, G, & B channels from facial region of interests. The authors of [53, 54] only predicted pulse rate and this work is extended in predicting pulse rate variability as well.

All the related work discussed above utilizes filtering and digital signal processing to extract PPG signal from the video which is further used to estimate the PR and PRV. The method proposed in [42] is person dependent since the weights will be different for people with different skin tone. In contrast, we propose a deep learning model to predict the PR and PRV which is independent of the person who is being trained. Thus, the model would work even if there is no prior training model built for that individual and hence, making our model robust.

Advances in computer vision and machine learning have the potential to address limitations of current ergonomics state-of-the-art methods for collecting force exertion data in workplace exposure assessments. Automated video exposure assessment has been used in previous studies to automatically quantify repetitive hand activity with the use of digital video processing [55]. Using video recordings to measure the ergonomic risk factor of repetitive motions, the investigators developed algorithms that tracked the hand and calculated kinematic variables of tasks such as frequency and speed [56]. The authors of [57] demonstrated that marker-less video tracking algorithm can be used to measure duty cycle and hand activity levels in repetitive manual tasks. The computer vision-based motion capture has been used previously to track and build on-site biomechanical model of the body and minimize work related ergonomic risk factors on construction sites [58–60]. The computer vision approach provides a promising tool for quantifying ergonomic risks from repetitive movements and potentially non-neutral postures; however, force exertion levels are another key ergonomic risk factor that computer vision techniques have not been developed to detect force exertion levels that may associate with injury risks.

Photoplethysmogram (PPG) is an optical technique for the volumetric measurement of the organ. It generates a pulsating wave based on the changes in volume of the blood flowing inside the arteries. Recently, there has been growing interest of the researchers in exploring PPG signal. Till date, the PPG signal has been used to extract information such as oxygen saturation level, blood pressure, respiration rate, pulse rate, and pulse rate variability. It is also a promising technique that is used in early screening of various atherosclerotic pathologies [61]. The amplitude of the PPG has been used as an indicator to vascular distensibility [62]. This information is used by anesthesiologists to judge subjectively whether a patient is sufficiently anesthetized for surgery. PPG waveform can be a useful tool for detecting and diagnosing cardiac arrhythmias as well. The researchers have also analyzed first and second derivative of PPG signal. The first derivative of the PPG can also be used to calculate the augmentation index which is a measure of arterial stiffness [61]. The measure of arterial

stiffness can be further be related to vascular aging [63]. There have been numerous applications of PPG signal and researchers are still exploring the potential of this signal.

3. DATA COLLECTION

This chapter focuses on the novel experimentation methodology developed for collecting the data to train a machine learning model. There are two different kind of experiments conducted to capture data related to cardiovascular parameters (pulse rate and pulse rate variability) and force exertion level of the person respectively. Both the experiments are conducted at Purdue campus under the Institutional Review Board protocol No. 1707019484 & 1708019605. The details of both the experiments are given below.

3.1 Experiments for cardiovascular parameters

The experiment is designed to collect data in order to predict pulse rate and pulse rate variability. From here on, this experiment would be denoted as Experiment A.

3.1.1 Set-Up

In order to predict the vital health metrics, the face video of the person are used. Therefore, the foremost task in the experiments is to capture the human facial videos. The videos are captured using a 5 MP front-facing Hello face-authentication camera (1080p HD) from Microsoft Surface Book, that has 30 frames per second (fps). It is a Microsoft-designed infrared camera with designated infrared light source to increase accuracy of face recognition. The camera is capable of capturing red, green and blue color bands. The authors of [39] suggested that green channel of the video outperforms blue and red channel in estimating the health parameters. Therefore, the green channel of the camera is utilized to maximize the signal noise ratio for

Fig. 3.1. Experimental set-up. The contact probe of Shimmer3 GSR+ is attached to the earlobe and laptop camera is placed around 0.5 m away from the subject.

model training. The features obtained from these videos will be used to predict the health metrics.

In order to train the data, true values of PR and PRV are calculated using a contact measurement device, Shimmer3 GSR+, that records the ground truth PPG signal. The earlobe is chosen as the suitable position for recording PPG since it is close to the face. This proximity is cared about since face is used for our video recordings as well. Subjects were asked to sit still for capturing the video, facing towards the camera at a distance of approximately 0.5m, and PPG signal was recorded simultaneously through Shimmer3 GSR+ device. One minute video was recorded for each subject. The longer the duration for PPG signal, the better estimate it gives for pulse rate variability [64]. Therefore, 1 minute was considered as the sufficient time for our

experiments. The set-up for conducting experiments is shown in figure 3.1 where a person is sitting still in front of the camera with Shimmer attached on the earlobe.

3.1.2 Experiments

The measurement for each subject is obtained in a room with constant lighting and background environment to reduce the variations in terms of illumination in the videos and across different subjects. The method of conducting the experiments is shown in Figure 3.1. The study involved people of different skin colors, race, and age to introduce variability in the data. Different activity levels are designed that need to be performed by each subject. The different activity levels at which the measurements were collected are: Rest Position, Brisk Walk, and Exercise, and are described next.These different activity level ensures variation in the ground truth values of pulse rate and pulse rate variability, thus providing variation in the dataset.

Activity Levels

The different activity levels designed are:

Rest Position: The first experiment was conducted when each participant was at rest condition. Each subject was asked to relax and sit still in front of the camera. The video and Shimmer device recordings were performed simultaneously. The main aim for this activity level was to capture the cardiovascular parameters at normal levels.

Brisk Walk: The next experiment involved data collection of the same participants after they were asked to do brisk walk for 0.25 mile at a speed of 3-4 mph on the treadmill. The video and shimmer recordings were captured as soon as the subject complete the brisk walk. It was made sure that there was no time gap between the walking and video recording. This helped in capturing the changed health parameters due to walking and helped obtain variety in pulse rate in the collected data set.

Exercise: The last experiment involved the task of performing either push-up or sit-up by each participant. All subjects were asked to perform as many push-ups or sit-ups as they can such that they exert themselves to their full capacity. The typical for this exercise was 10-25 push-ups or sit-ups depending on the individual. The video and Shimmer device recordings were performed as soon as each participant completed the task of performing this exercise. This activity was designed to elicit high pulse rate since the individual was working out at their full capacity.

The reason for three different kind of activities was to capture wide range of conditions for the training data. During the rest position, the pulse rate of the subject will be normal but it will increase as person switches to brisk walk and then to exercise. Figure 3.2 shows the difference in the mean value of pulse rate (bpm) for different subjects under three levels of activity. The mean pulse rate for rest, walk, and exercise condition was recorded to be 72.9, 79.6, and 98.53 respectively. Each subject was asked to quickly record the video and PPG after each activity so that the pulse rate does not go back to normal before taking the video. Each subject was given a rest of 10-15 minutes before each activity so as to recover the pulse rate and bring it back to normal level. The rest time was enough to recover as for healthy adults, pulse rate falls at a rate of about 20 beats per minute after intense exercise [65]

3.2 Experiments for exertion level

An another experiment has been designed to collect the data that will be used to predict force exertion level. A study was conducted where each subject exerted varying levels of muscle force. During these activities, we collected the videos of the person performing the activity as well as data regarding the volumetric blood flow by capturing PPG signal using pulse oximeter. Figure 3.3 shows the complete set-up we used in our study. From here onwards, this experiment would be called as Experiment B.
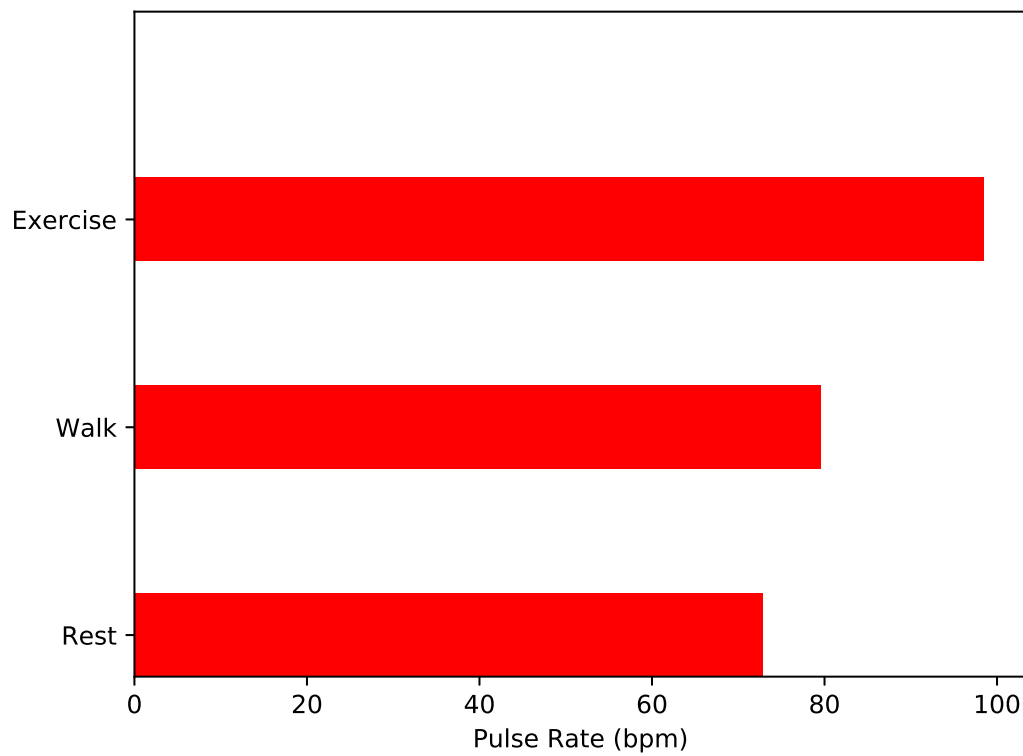
Fig. 3.2. The mean value of pulse rate measured in beats per minute for all the subjects under different activity levels.

Fig. 3.3. The experimental setup with a subject holding a grip dynamometer and pulse oximeter attached to the earlobe. The GoPro attached on the tripod is used to capture the video

3.2.1 Study participants

Twenty healthy volunteers participated in this study. The participants were recruited from a university population through email including a description of the study. This study was reviewed by the university's Institutional Review Board and all participants provided informed consent. The only exclusion criteria were current injuries that prevented participants from performing force exertions. Sixteen males and 4 females participated in the study, all were right hand dominant, and their ages ranged from 18 to 29 years. The details of all the subjects that participated in the study is given in Table 3.1.

3.2.2 Study Setup

The power grip dynamometer was used to measure the grip force of each subject. This devise helps in measuring the maximum isometric strength of the hand and forearm muscles and hence helps us collecting the ground truth of force exertion level for each subject.

A GoPro camera was used to capture the video of our subjects while they were performing different kind of activities. We placed GoPro in front of the subject, around 0.5 meter away from face, and video recorded the subject during the entire experiment. It is a 12 MP camera and recordings are done at 50 frames per second.

The photoplethysmogram (PPG) signals were recorded using using pulse oximeter. The PPG signals were captured by Shimmer GST+. This device has a contact probe that is attached to the earlobe. The earlobe is chosen as the suitable position for recording the PPG.. Although the signals could be estimated using the non-contact methodologies [42], in this study the actual PPG signals were recorded using pulse oximeter to minimize the errors of estimation.

Table 3.1.
Data for 20 subjects in Experiment B

	Female (n=4)		
	Mean ± SD	Min	Max
Age (years)	20.0±1.4	19	22
Weight (lb)	124.0±33.9	100	148
Grip Force (lb)	47.0±15.4	30	62
	Male (n=16)		
	Mean ± SD	Min	Max
Age (years)	20.8±2.7	18	29
Weight (lb)	133.8±21.7	110	168
Grip MVC (lb)	88.8±20.4	62	118

3.2.3 Study Design

At the beginning of the data collection session, participants were provided a description of the study, and written consent was collected. First, the subjects were seated in front of the white background to minimize the noise in video processing in detecting the face. The handheld dynamometer was calibrated as per the hand size to ensure standardized and comfortable gripping postures for each subject. This follows attaching pulse oximeter's contact probe properly to the subject's earlobe.

Participants were given a 5-minutes practice period to familiarize with the devices and environment. The overall study involved three different activity levels at different setting of grip dynamometer. In the first trial, each participant performed a grip exertion at maximum force that they are capable of. The subjects were instructed to maintain the maximum force for 9 seconds (note that although the magnitude of the force may decrease during the 9-seconds, participants continued to exert their maximum effort). The recordings were stopped after 9 seconds.The second exertion trial was 0% grip force. In this trial, subjects were asked to hold the grip dynamometer without exerting any grip force. The subjects rested for 5-10 minutes between each force exertion levels to prevent fatigue effects from carrying over to the next force exertion trial. The rest period was also enough for subjects to recover from increased pulse rate during the activity [65]. Finally, the last trial was force exertions at 50% of maximum force . In this trial, each subject was asked to exert exactly 50% of their maximum grip contraction. The distribution of the grip force for different subjects is reported in Table 3.1.

4. METHODOLOGY

This chapter gives a detailed explanation about the methodology used to process the recorded data during our experiments and convert it into meaningful features that can help us in predicting the cardiovascular parameters (pulse rate, pulse rate variability) and force exertion level for each subject.

The overall method adopted for processing data from both the experiments includes three main steps:

1) Processing the videos using techniques of computer vision to extract the relevant regions from each frame of the video

2) Feature extraction from each of the processed frame.

3) Training a neural network model that will take extracted features as an input and outputs the predictions.

4.1   Video Processing

In this work, videos are processed using an existing state-of-the-art algorithm called Deepface [47]. This algorithm has been developed by the researchers at Facebook. Deepface is a face recognition algorithm that consists of four main stages: 1. Detect 2. Align 3. Represent, and 4. Classify.

There have been other work in developing algorithm for facial recognition [66–71], but Deepface [47] reached an accuracy of 97.35% in Labeled Faces in the Wild (LFW) dataset and reduced the error in face recognition of current state-of-the art by more than 27%. The high accuracy in Deepface is achieved by revisiting both alignment and representation step. 3D face alignment has been done using piecewise affine transformation and face representation is derived using 9-layer neural network which

is a key for the high performance. Therefore, we utilized Deepface for recognizing faces in our approach.

4.1.1 Video Processing for Experiment A

The video is recorded for 1 minute for each subject under different activities as referred in 3.1.2 but for our analysis, first 50 seconds of the video are used (The last 8-9 seconds video was corrupted for 3 subjects and in order to keep consistency, we trimmed our videos to 50 seconds). Entire video is broken into frames where each frame is composed of red, green, and blue color bands. The first step in the video processing involved the detection of human face in each frame of the video. The detected human face is then aligned automatically by DeepFace using 3D alignment method [47]. The aligned face was cropped from the image using the landmark points on the face shown in Figure 4.2.

4.1.2 Video Processing for Experiment B

The 9 seconds video of each subject is trimmed to 7 seconds before passing it to Deepface. The first 2 seconds of videos are removed because each subject requires initial 1-2sec to reach to the required force level. Each video is recorded at 50 frames per second and hence, consists of 350 frames We process all these frames using Deepface that recognizes and aligns the face of each subject across the frames using 68 landmark points on the face. Figure 4.1 shows how deepface is used to extract faces from the each frame in the video. Figure 4.1 (a) is an example of an actual frame in the video. DeepFace recognizes the face of the person in each and crops the face out of it as shown in Figure 4.1 (b). This algorithm helps identify 68 landmark points on the face as depicted in Figure 4.1 (c) and track these 68 landmark points over the whole video The 68 landmark points represents the contour of the face, eyebrows, eyes, lips, and nose. Detecting and aligning the face in each frame of the video is

Fig. 4.1. The steps followed for feature extraction from each frame of the video. (a) The actual image (one of the many frame) from the video captured during the experiment. (b) The detected and aligned face using DeepFace. (c) The face along with the 68 landmarks on it. These 68 landmark points are used by DeepFace in face recognition.

one of the most critical step in our overall methodology, because relevant features to train a neural network will be extracted from the output of Deepface.

4.2 Feature Extraction

The extraction of "right" features is important as it plays significant role in training a neural network. The choice of relevant features leads to the simplification of the models which in turn requires shorter training time [72]. "Right" set of features helps in avoiding the curse of dimensionality and leads to generalization of the model by reducing the variance in the model [73].

4.2.1 Feature Extraction for Experiment A

The image was cropped so as to have only face feature and removing extra pixel values from the images. The forehead was carefully retained since it contains the maximum information about blood perfusion inside the arteries [42]. Therefore, the

amount of information contained in the pixel values of forehead was important for our analysis. The cropped images were used for training the deep learning model. The stages of video processing are shown in Figure 4.2. The pixel instensity values becomes the relevant set of features, D_0, from experiment A.

4.2.2   Feature Extraction for Experiment B

Features from Videos

Deepface utilizes the information of 68 landmark points on the face. The proposed method use 128 landmark points on the face as shown in Figure 4.4. Based on 68 landmark points, we locate 60 more landmarks on the face that lies on the left and right cheeks. Thirty landmarks on each cheek is located based on the location of landmarks on the contour of the face and eyes. Different landmark points can be grouped together based on the location on the face as: 1: Contour of Face (17 landmarks), 2: Left Eye with left eye brow (11 landmarks), 3: Right eye with right eyebrow (11 landmarks), 4: Nose (9 landmarks), 5: Lips (16 landmarks), 6: Left Cheek (32 landmarks), 7: Right Cheek (32 landmarks).

All the 128 landmark points are tracked in 350 frames for each video. The location (x and y co-ordinate values) of each landmark is extracted and based on the location, the average movement of each landmark with respect to its location in the first frame is calculated over the entire video. For each video, average movement, d_j,of each landmark, j, is given in equation 4.1

$$d_j = \frac{\sum_{i=1}^{n} \sqrt{(x_{ji} - x_{j1})^2 + (y_{ji} - y_{j1})^2}}{n} \tag{4.1}$$

where n is the number of frames in the video. Thus, the set $D_1 = \{d_1, d_2 ....d_{128}\}$ is our first set of features used in the prediction of exertion level. These features are potential indicator of exertion level as they depict how each point on the face moves in the entire video when subjects are asked to perform different exertion level activity.
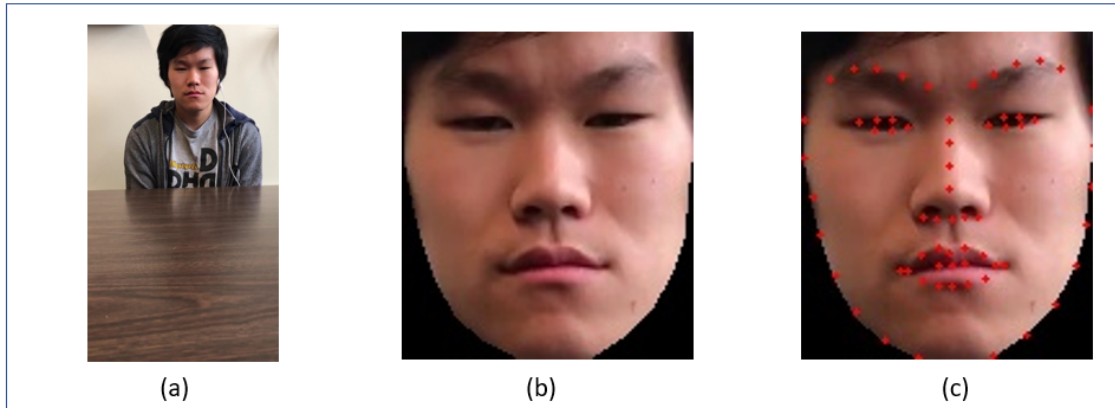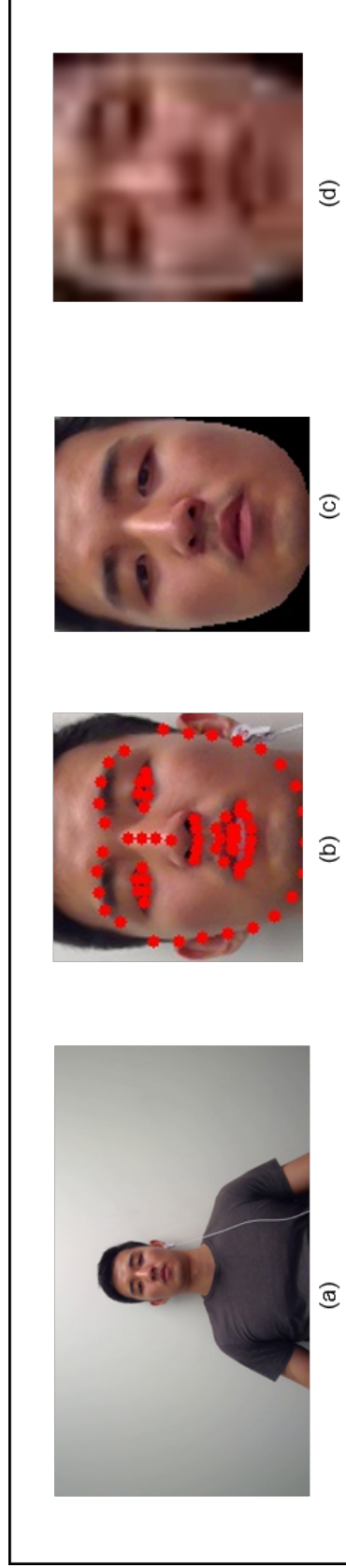
Fig. 4.2. The steps followed for feature extraction from each frame of the video. (a) The actual image (one of the many frame) from the video captured during the experiment. (b) The detected and aligned face using DeepFace along with landmark points. (c) The face is cropped using landmark points so as to get only required face features. (d) Each frame is downsampled to 20x20 image

Features from PPG

The other set of features is derived using PPG signal captured during our data collection experiments. PPG signal of each subject consists of multiple beats where each beat is defined as the set of consecutive values of PPG having a maximum PPG value between two minimum values as highlighted in yellow color in Figure 4.3 (a). Therefore, for each beat we have a starting point denoted at T1, maximum point denoted as T2, and end point denoted at T3. The total number of beats of 7 seconds recorded PPG were varied from 8 to 12 beats. Therefore, from each PPG signal following features are extracted:

- Time interval between T1 and T2 is extracted for first 5 beats

- Time interval between T2 and T3 for first 4 beats

- Time interval between T1 and T3 for first 4 beats

- Standard deviation of PPG values at T2 for all beats

- Standard deviation of PPG values at T1 for all beats

- Mean of three time intervals: T1 and T2, T2 and T3 and T1 and T3

- Standard deviation of three time intervals: T1 and T2, T2 and T3 and T1 and T3

The above mentioned 21 features were extracted from PPG signal corresponding to each video. The set of these 21 features is referred to as D_2 in rest of the document. The PPG features extracted from the signal corresponds to the cardiovascular activities of the person during different experiment levels. In this work, we extracted these features from PPG signal that is captured using a contact device. Recently there have been advancements in passively estimating the PPG signal using facial video without the need of any contact device. There are many state-of-the-art techniques discussed in [39–42, 48–51] that utilizes videos of the human to extract PPG signal.
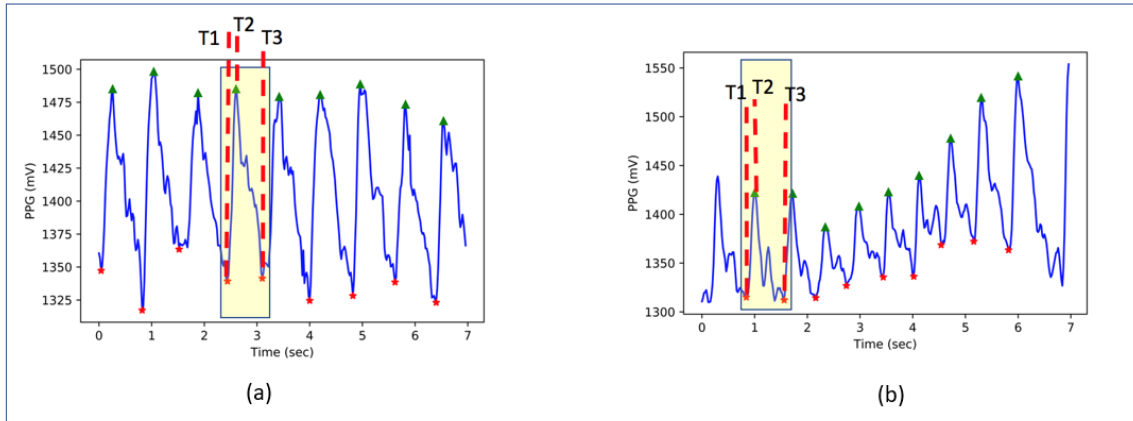
Fig. 4.3. Collected PPG Signals (a) 0% grip force (b) 50% grip force;
T1: time at first local minimum of the beat, T2: time at local max
of the beat, T3: time at the end of the beat
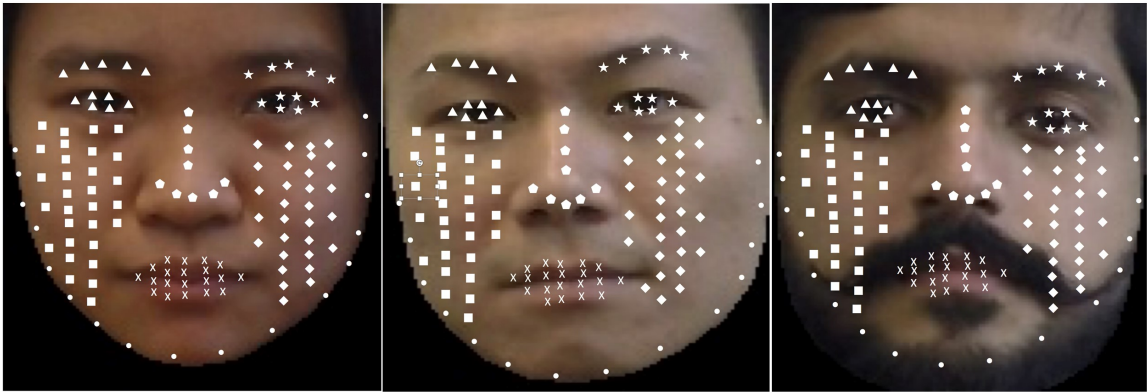
Fig. 4.4. The location of 128 landmark points on the face for different
subjects. Additional 60 landmarks have been identified on the face
for efficient model training

The features extracted for model training is our main novelty as none of the other
authors have extracted such features and used machine learning to predict the force
exertion level. This is the first ever work that utilizes such facial and PPG features.

$x_1 \longrightarrow$

$x_2 \longrightarrow$

$x_3 \longrightarrow$

$x_n \longrightarrow$

Input Values **Input Layer** **Hidden Layer 1** **Hidden Layer 2** **Hidden Layer 3** **Output Layer**

Fig. 4.5. The architecture of a fully connected neural network with three hidden layers for training features from Experiment A

4.3 Model Training

4.3.1 Model Training for Experiment A

We train a deep learning model using TensorFlow to estimate the health metrics. The model was trained through multi layered neural network. The architecture of the neural network, NN_0, used for training D_0 is shown in figure 4.6. A fully connected neural network with three hidden layers is used. The network consisted of rectified linear units (ReLU) and the rectifier activation is given as $f(x) = max(0, x)$, where x is the input to the neuron. The network is trained using back propagation algorithm with mean squared error as the loss function. The green color band was shown to be the best source for extracting information about the health parameters [39]. Hence, we utilized all the pixel values corresponding to the green channel of each frame to train our machine learning model. Image from each frame is down sampled to 20x20 image. Hence, we used 400 features from each frame to train the model. The features were normalized to bring them in a range of [0,1] so that it becomes easy for neural network to learn from the data. The down sampling was done to reduce the computational expense of our mo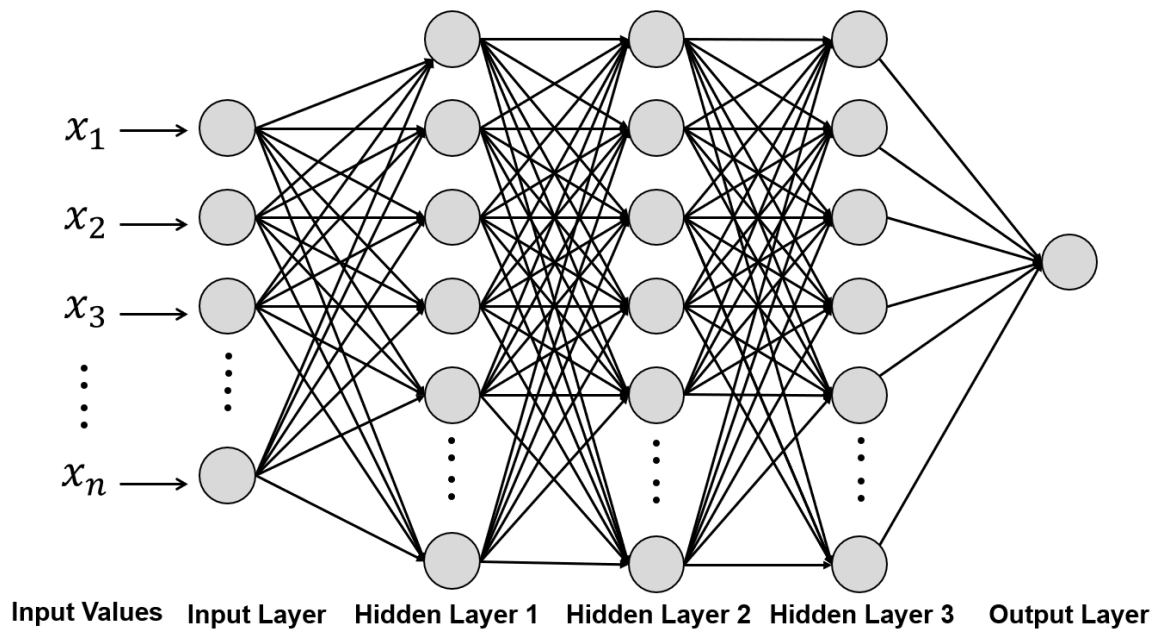del. The actual response value i.e. PR and PRV was extracted from the PPG signal recorded during our experiment. In order to extract actual PR from PPG, we computed power spectral density (PSD) of the PPG signal using fast Fourier transform (FFT) algorithm. The PR is then estimated as the frequency corresponding to the maximum power in the PSD (PR= 60. f bpm), where f is the required frequency. The LF and HF components of PRV are calculated by computing the area under the PSD curve between specific frequency range. For LF, the frequency range is 0.04-0.15Hz and for HF, it is 0.15-0.4Hz. We used PR, HF, and LF as a response variable to train our model.

The PR and the two components of PRV estimated from PPG were trained for the whole video of 50s which contains 1500 frames. The computation accuracy of PRV increases with video duration [64], therefore, we utilized the whole video at once to train the network.

4.3.2 Model Training for Experiment B

After all the relevant features D_1 and D_2 are extracted, we train two neural networks : NN_1 & NN_2 to predict the three levels of exertion level of human. The feature set D_1 is used for training NN_1 that classifies 100% force exertion level and rest of other levels (0% & 50%) and further feature set D_2 is trained on NN_2 to classify between 0% and 50% level.

The architecture of both NN_1 and NN_2 used to train the features is same as shown in Figure 4.6. The extracted features D_1 & D_2 were used as the input data into a neural network with 1 input , 3 hidden and 1 output layers as shown in Figure 4.6. For each hidden layer, 35 neurons are used. The activation function used in the training of network is exponential linear units ($ELUs$) [74] as defined in equation 4.2. Batch normalization is used in each hidden layer [75]. The use of drop-out is one of the simplest way to avoid over-fitting of the neural network [76]. Drop-out rate was set to 50% to avoid over-fitting in all the three hidden layers. This will help in better generalizing the network for unseen data. In the output layer, two neurons were used for 100% and rest (0% and 50%) for NN_1 & 0% force exertion and 50% force exertion level for NN_2. The best performance of the network is achieved with using Adam [77] as an optimizer along with categorical cross-entropy as a loss function.

$$f(x) = \begin{cases} x, & \text{if x} \geq 0 \\ \alpha \times (e^x - 1) & \text{if x} \leq 0 \end{cases} \tag{4.2}$$
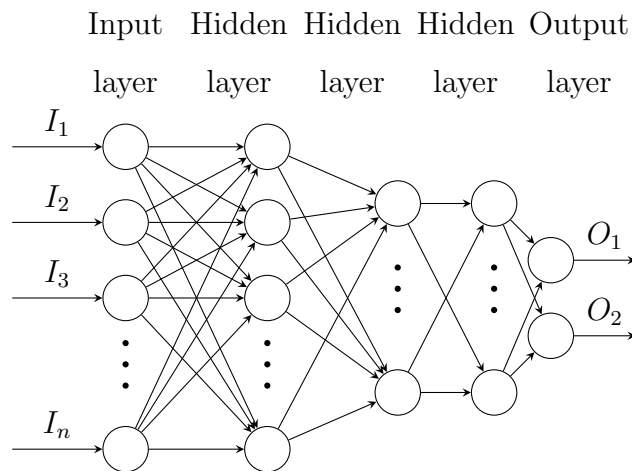
Fig. 4.6. The architecture of a fully connected neural network with three hidden layers

5. RESULTS

This chapter details about the prediction results in case of both the types of problems considered in this work: regression problem in order to predict pulse rate, and pulse rate variability; classification of prediction of force exertion level of the human. The results are computed from the different neural network models. The point to note here is that, for regression problem, the output layer had 1 neuron with mean squared error and a loss function, whereas in classification problem, the output layer of the neural networks used had two neurons and categorical cross entropy has been used as the loss function.

5.1 Result for predicting Pulse rate and pulse rate variability

5.1.1 Pulse Rate

The values of PR predicted by the model were compared with the true values calculated from the readings of contact device and the errors were calculated accordingly. The mean absolute percentage errors were calculated using leave-one-out cross validation. To be more specific, since there are 20 subjects in total, nineteen out of the twenty subjects are chosen and picked out all observations from those nineteen subjects to make up our entire training set. All observations for the subject that was left out from the training set were considered as the test set. This procedure is iterated for each subject to make sure that model is tested on each individual. Since the data from test set is completely new compared to the training set, this tells us how the model predicts on subjects it has never seen before, regardless of skin tone, race, and facial features.

The Mean absolute percentage error (MAPE) and Root Mean Squared Error (RMSE) for our predictions is calculated. The mean of errors on all 20 subjects was found to be 4.6%. Similarly, RMSE value for our test set is found out to be 4.39. The authors of [53] reported a RMSE of 9.52 on test set in predicting PR meaning that our model outperforms theirs and shows a reduction in RMSE by 53% for predicting PR.

Figure 5.1 shows how the test and train loss varies with number of iterations run by the neural network. For computation, mean squared error is used as the loss function. The number of iterations were chosen based on the behavior of test and train loss. If the number of iterations were too low, it lead to under-fitting wherein both train and test error were high and if number of iterations were too many, it lead to over-fitting. In order to avoid these scenarios, the model is run for 170 epochs.

Figure 5.2 shows a scatter plot between predicted and actual values of pulse rate. The straight line shown is a 45 degree line ($y = x$), and the closeness of the scatter points to the straight line indicates the high accuracy of our model.

5.1.2 Pulse Rate Variability

Another health metric considered in this paper is pulse rate variability. The model is trained using High-Frequency component and Low-Frequency component of PRV. The separate models to train our model to predict the HF component and the LF component are designed.

The models are tested using leave-one-out cross validation method similar to the pulse rate. The model makes predictions on the user who has not been seen before in the training data. The number of iterations to run our PRV model is chosen as 200. The test and training losses with iterations for the LF and the HF component of the PRV are depicted in Figure 5.3 and Figure 5.5, respectively.

For predicting normalized LF, mean absolute percentage error on test set is 4.58%, and the root mean squared error is 3.49. On the other hand, the MAPE for normalized
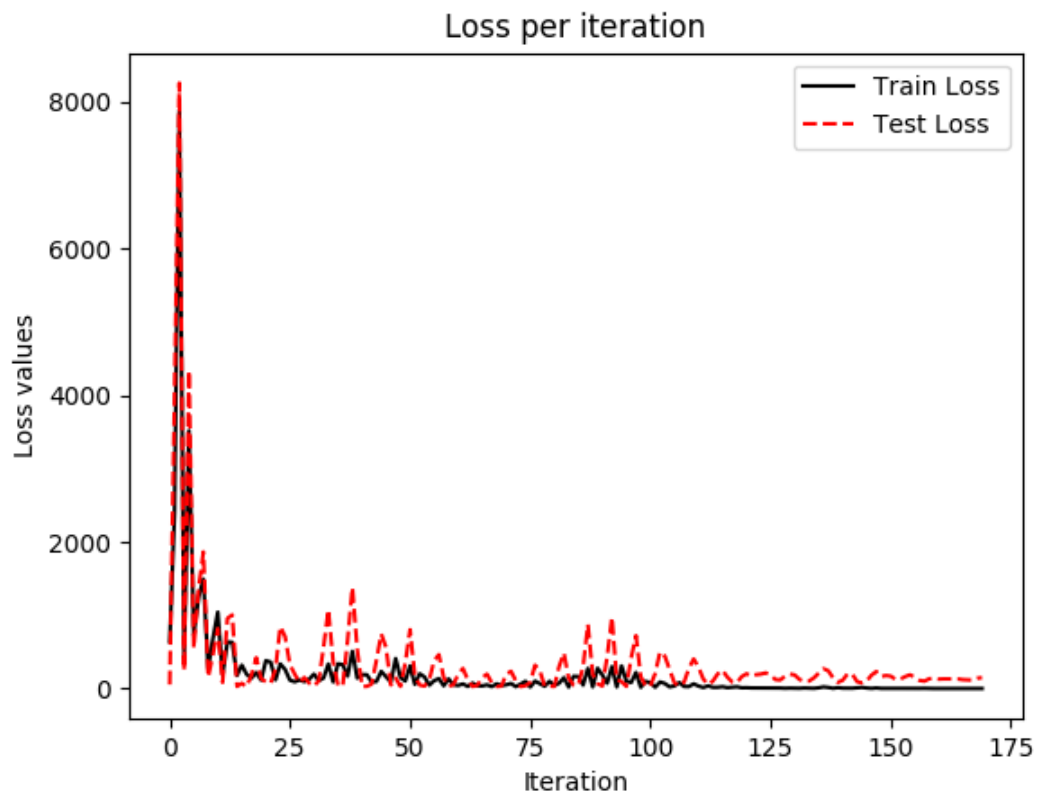
Fig. 5.1. Behavior of train loss and test loss for predicting pulse rate.
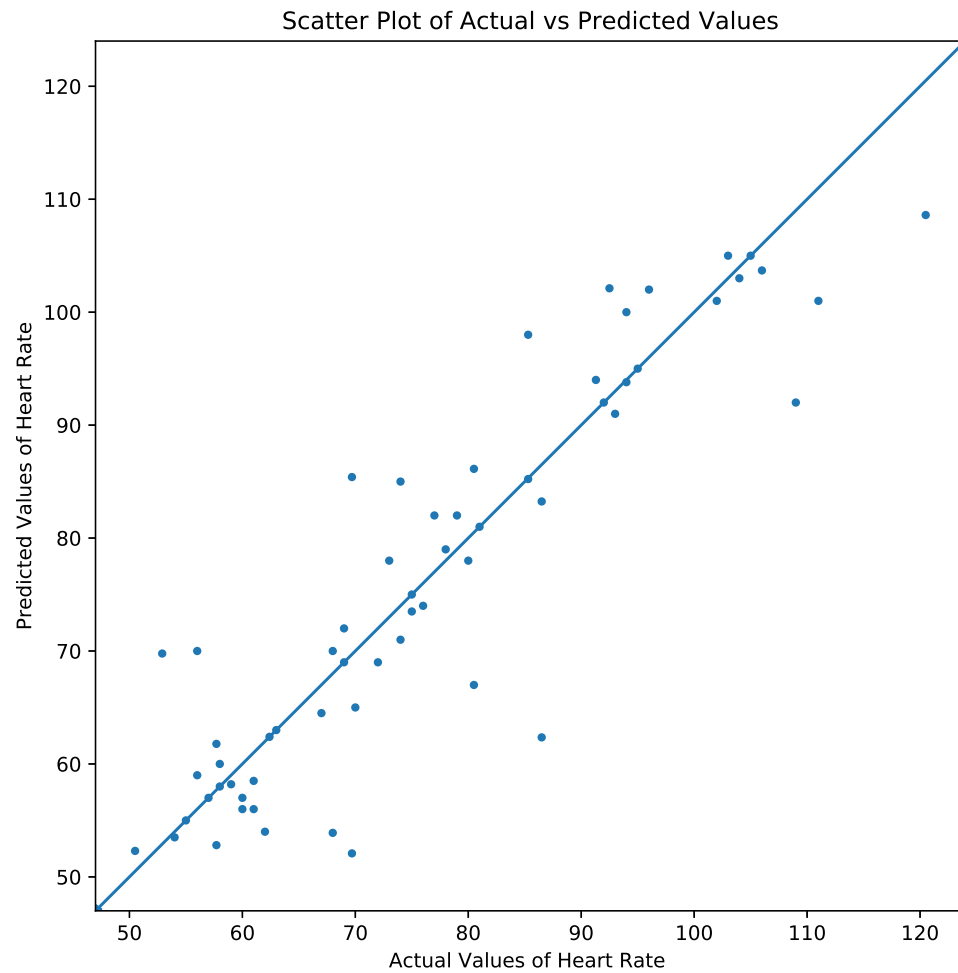
Fig. 5.2. Scatter plot of the predicted PR value vs. the ground truth PR value. The straight line is function $y = x$. The closeness of the points to the line indicates the model accuracy.
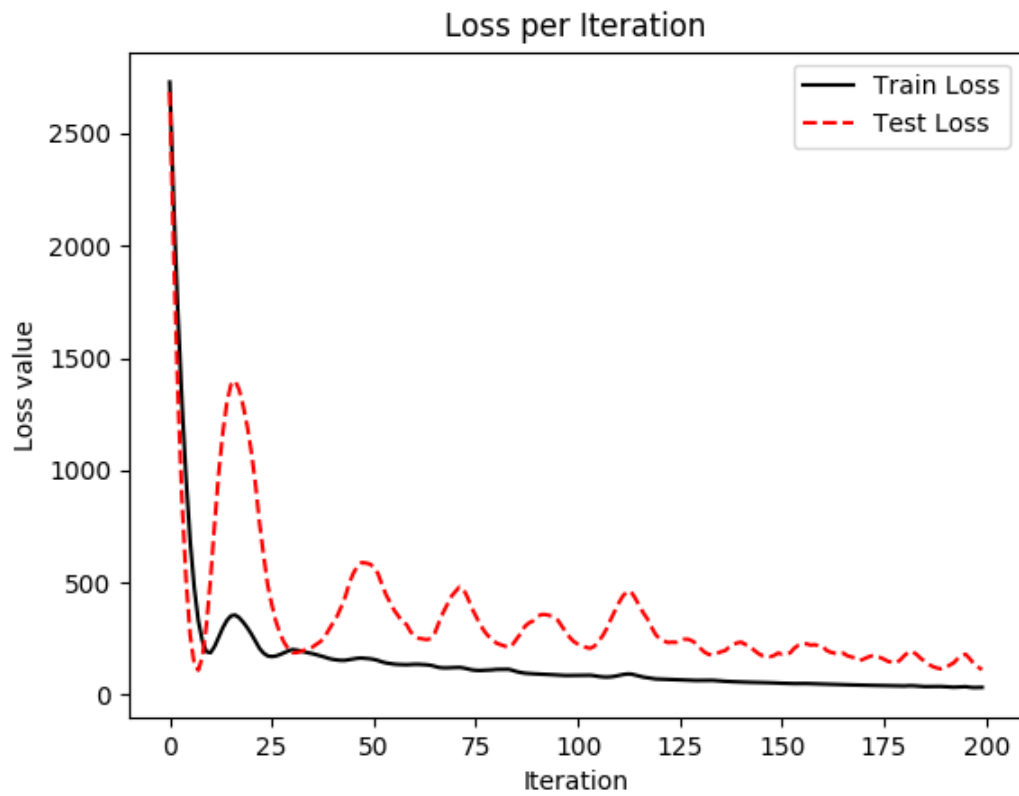
Fig. 5.3. Training loss and test loss for predicting Low Frequency (LF) component of PRV
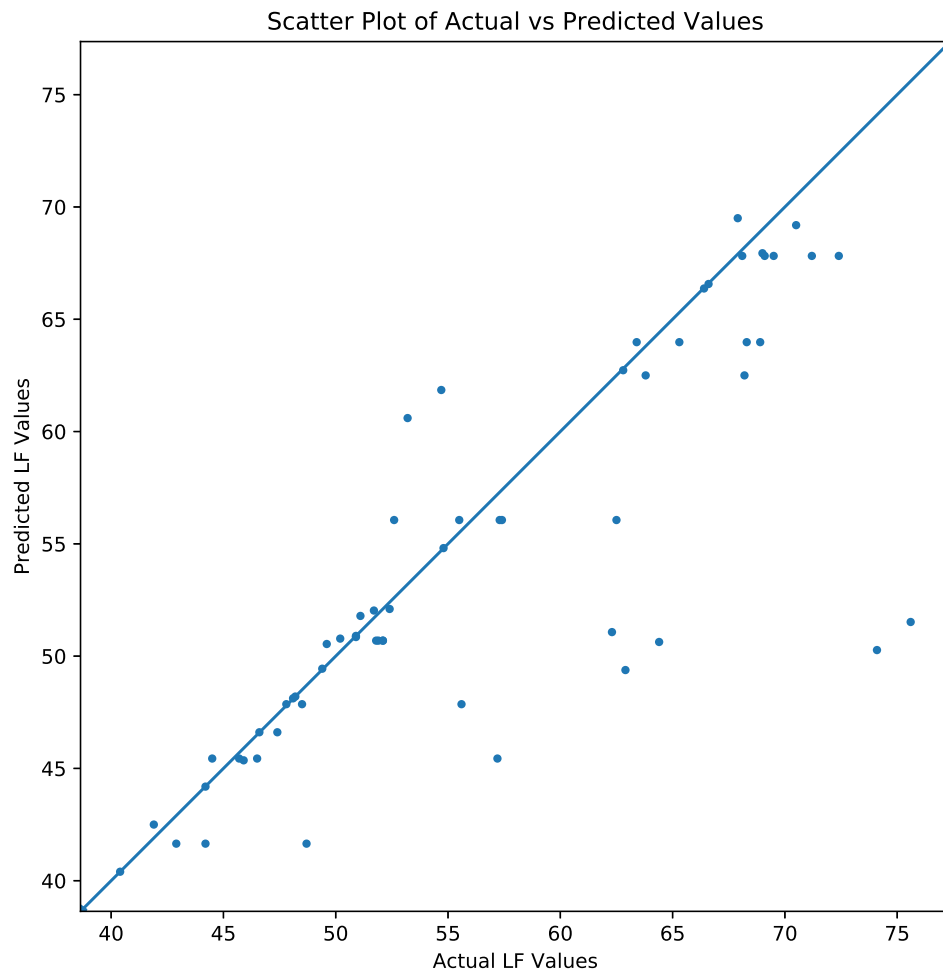
Fig. 5.4. Scatter plot of the predicted LF value vs. the ground truth LF value.
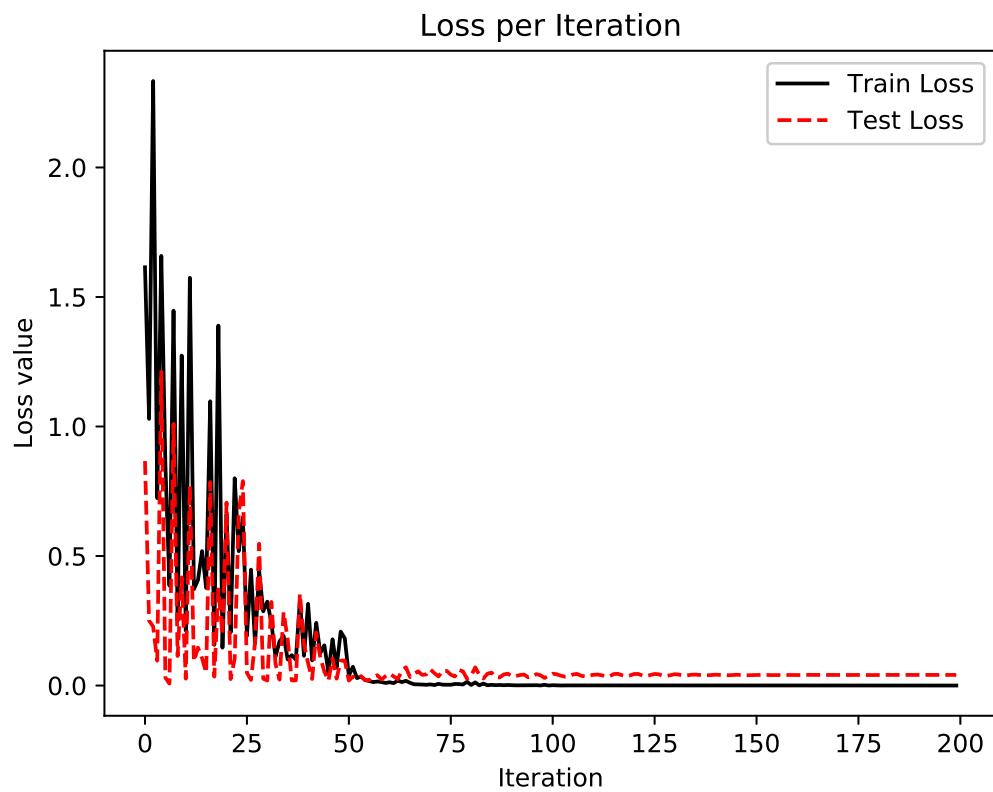
Fig. 5.5. Training loss and test loss for predicting High Frequency (HF) component of PRV.
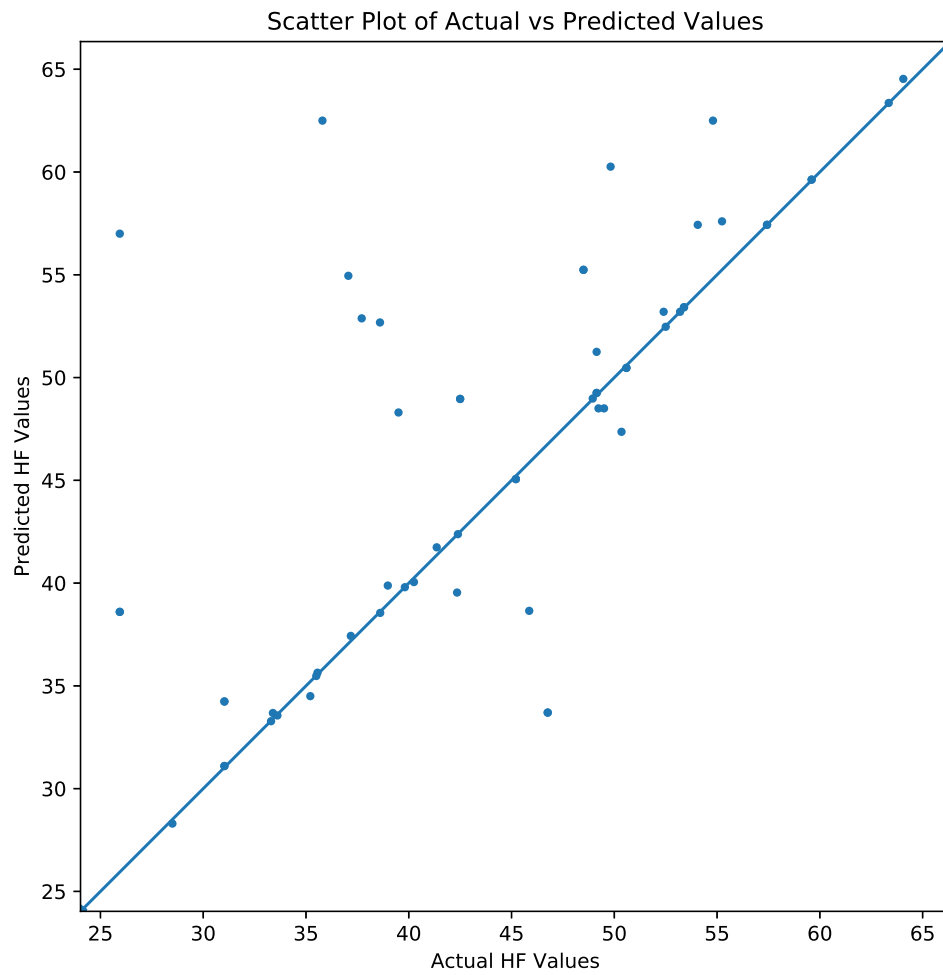
Fig. 5.6. Scatter plot of the predicted HF value vs. the ground truth HF value.
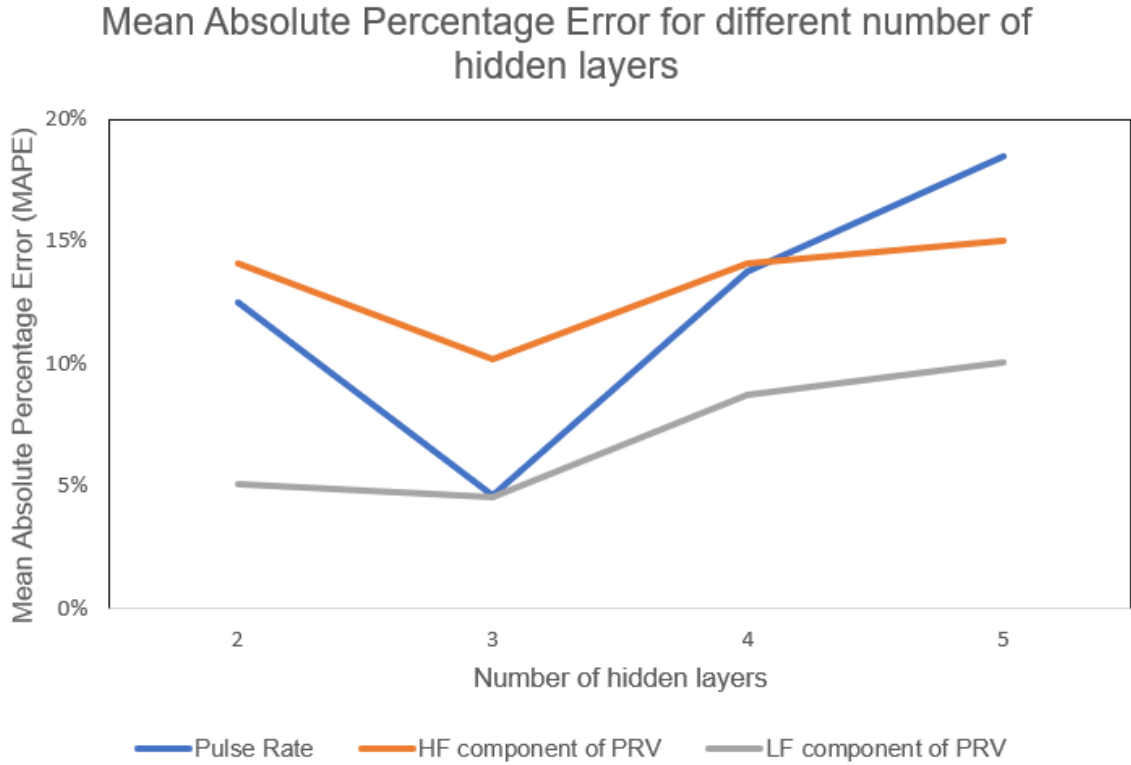
Fig. 5.7. Plot of behavior of Mean Absolute Percentage Error against number of hidden layers in a neural network for predicting cardiovascular parameters

HF is found to be 10.2%, and the RMSE is 4.96 for test data. The mean RMSE of PRV for our model is 4.3 whereas the mean RMSE taken over different skin colored people in [42] is 25.3 thus providing 83% decrease in the RMSE for predicting PRV. Figure 5.4 and Figure 5.6 depict the comparison between the actual values and the predicted values for the two components of the PRV, respectively.

The neural network used in predicting all the above stated cardiovascular parameters uses three hidden layers in their architecture. Figure 5.7 provides the reasoning for this choice. Figure 5.7 shows how MAPE changes with respect to the number of hidden layers in the neural network while predicting different cardiovascular parameters. The value of MAPE is higher when 2 hidden layers are used as compared to 3 hidden layers. This could be because of insufficient number of layers to learn the

pattern in the data well. Also, we see that as the number of hidden layers exceeds 3, there is increase in the value of MAPE for all the three variables of interest. This could be because network might be over-fitting the training dataset because of higher number of layers and hence doesn't generalize well on the test dataset, and hence higher MAPE.

5.2 Results for predicting Force Exertion Level

This section summarizes the classification results of both neural networks i.e., NN_1 & NN_2. The performance of final model is discussed along with the noise analysis that shows the robustness of the trained model.

5.2.1 Force level classification using D_1

The classification between 100% (group A) and 0% & 50% (group B) is done using NN_1. The model is trained using a fully connected neural network as architecture that utilizes average movement of landmark points. The neural network is trained for 200 epochs. The number of epochs are decided based on the performance of test and train loss curves . The test loss for the performance of neural network is reported using leave one out cross validation approach meaning during training the neural network, the data for average movements of all force level for 19 subjects is used and once the model is trained, the performance is measured using data from 1 subject that has been left out of training. This approach has been repeated for all the subjects and the average accuracy on test set has found to be 90% . The graph in figure 5.8 represents the behavior of accuracy and loss value for test and training dataset for one of the subject. Table 5.1 shows the accuracy results for all subjects. The subjects with 'o' represents that they have been correctly classified for a particular group and subjects having 'x' corresponding to a particular group shows that the class is predicted incorrectly. We can predict Group B correctly for all the
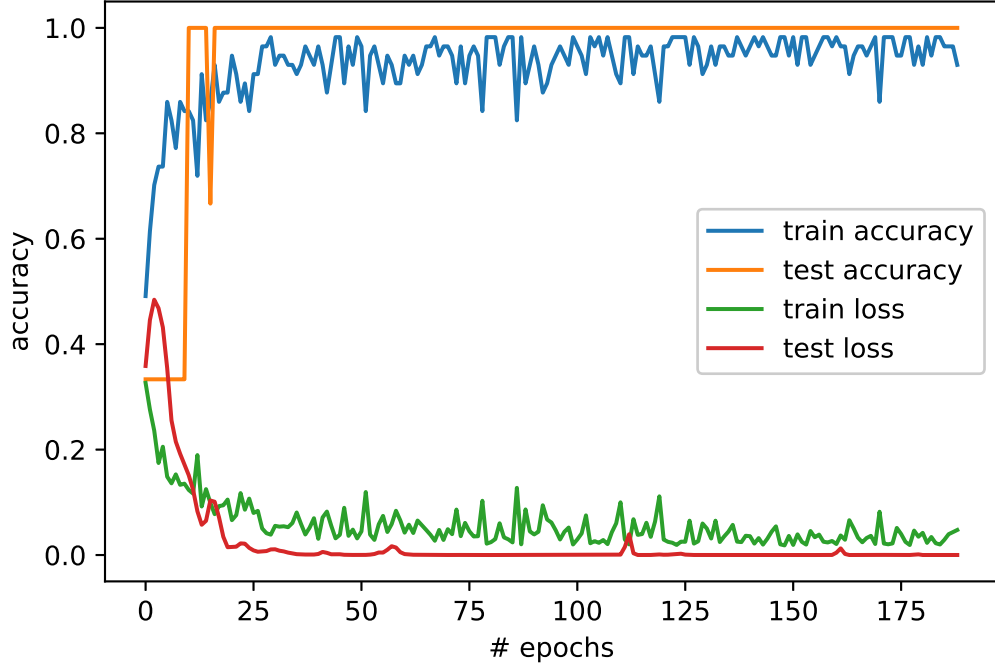
Fig. 5.8. The behavior of accuracy and loss values of NN_1 for test and train dataset against number of epochs

subjects except subject no. 6 and also we predict group A correctly 17 out of 20 subjects leading to an overall accuracy of 90%.

5.2.2 Force level classification using D_2

After classifying Group A and Group B using NN_1, NN_2 is used to classify the 0% and 50% force exertion level in Group B. This neural network model utilized all the features extracted from PPG that has been described in section 4.2.2. This neural network is trained for 175 epochs. The number of epochs are chosen such that model doesnot ovefit. The technique of early stopping [78] is used here to avoid overfitting, reduce variance in the model and generalize model well over unseen data. This model also uses same approach of "leave one out" as has been discussed in previous

Table 5.1.
Table showing prediction results for NN_1

Subject ID	1	2	3	4	5	6	7	8	9	10
Group A	o	o	o	o	o	o	o	o	o	o
Group B	o	o	o	o	o	x	o	o	o	o
Subject ID	11	12	13	14	15	16	17	18	19	20
Group A	o	o	x	x	x	o	o	o	o	o
Group B	o	o	o	o	o	o	o	o	o	o

Table 5.2.
Table showing prediction results for NN_2

Subject ID	1	2	3	4	5	6	7	8	9	10
Group B-0%	o	o	o	x	o	x	o	o	o	o
Group B-50%	o	o	x	o	o	o	x	o	x	o
Subject ID	11	12	13	14	15	16	17	18	19	20
Group B-0%	o	o	o	o	o	o	o	o	o	o
Group B-50%	o	o	x	x	o	o	o	x	o	o

subsection. The average accuracy on 20 subjects is 80% for NN_2. The behavior of the accuracy and loss values for testing and training data while training the neural network is shown in Figure 5.9. Table 5.2 shows the average accuracy results for each subjects. The subjects with 'o' represents that they have been correctly classified for a particular group and subjects having 'x' corresponding to a particular group shows that the class is predicted incorrectly. The model can correctly predict 0% force exertion level for 18 out of 20 subjects and 50% force exertion level for 14 subjects out of total of 20 leading to an overall accuracy of 80%.
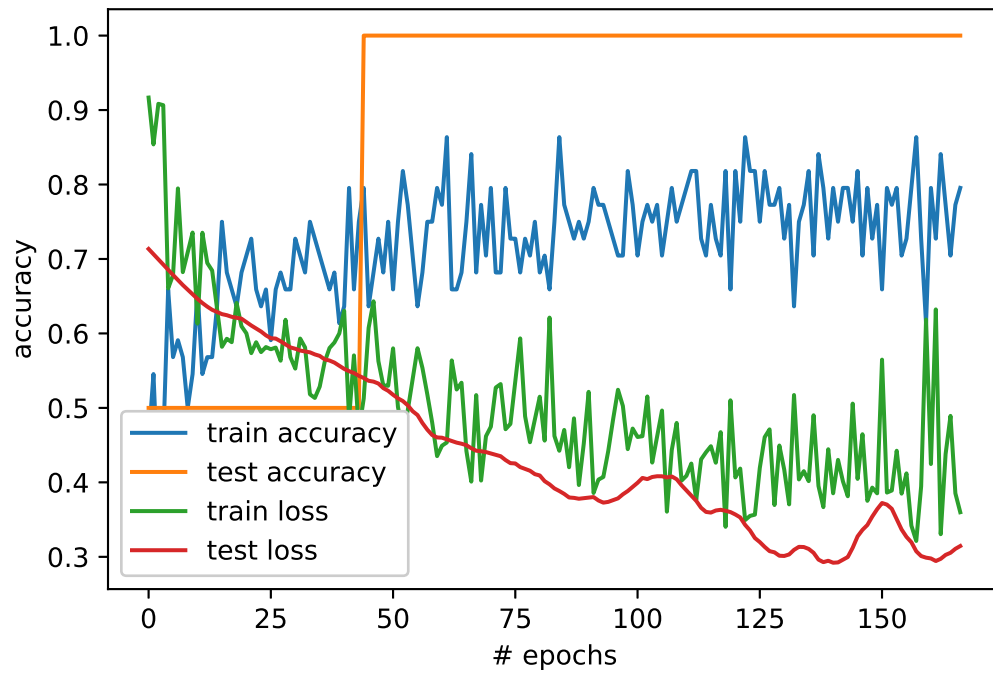
Fig. 5.9. The behavior of accuracy and loss values of NN_2 for test and train dataset against number of epochs

Table 5.3.

Table showing final prediction results from NN_1 & NN_2

Subject ID	1	2	3	4	5	6	7	8	9	10
0% force exertion	o	o	o	o	o	x	o	o	o	o
50% force exertion	o	o	x	x	o	o	x	o	x	o
100% force exertion	o	o	o	o	o	o	o	o	o	o
Subject ID	11	12	13	14	15	16	17	18	19	20
0% force exertion	o	o	o	o	o	o	o	o	o	o
50% force exertion	o	o	x	x	o	o	o	x	o	o
100% force exertion	o	o	x	x	o	o	o	x	o	o

5.2.3   Final Model

The proposed methodology breaks the process of classifying different force exertion levels into two steps and the accuracy results for two scenarios are presented separately. The overall accuracy of the two models combined is calculated to be 81.7%. The predictions for all the subjects has been combined together from NN_1 and NN_2 and has been shown in Table 5.3. There were some subjects that had same facial expression during both 50% and 0% force exertion level experiment and there was no significant difference between average movement of the 128 landmarks between these two levels. Hence, it is not advisable to use average movement of the landmark points as classifying feature between 0% & 50% force exertion level. Therefore, PPG features were extracted that related the cardiovascular parameters with force exertion level.

In this work, it was attempted to train one neural network that uses both facial features and PPG features together and classify three levels of force exertion. The accuracy result for such neural network was below 70% and the the further reason to train two different neural networks are given in section 6.2

5.2.4   Model Robustness

In order to check the robustness of the model that classifies force exertion level, we collected the data of 7 additional subjects while they were performing different kind of activity. The videos of the subjects using the same experimental set-up are captured when they were not performing any sort of force exertion activity, but are talking. Each subject was asked to speak a paragraph on themselves for 9 seconds for and their video along with PPG data was recorded during this activity. This activity is named as "talking".

Using the same processing technique, D_1 and D_2 set of features are extracted for all the subjects during talking. The set D_1 is passed through trained NN_1 and set D_2 is passed through trained NN_2 and predictions are made as of what activity level they belong to.

It is interesting to note that for all the 7 subjects when D_1 is passed through NN_1, it always predicts group B for the activity level which means that the algorithm is able to differentiate between talking and 100% and therefore gives high probability to group B. When set of features derived from PPG are used as an input to the trained NN_2, for 5 out of 7 subjects, the model predicted it to be 0% force exertion level and for 2 out of 7 subjects, model predicts as if subjects are at their 50% force exertion level. Note that the data corresponding to talking was not used while training the network. It is completely unseen data for our two trained neural networks.

6. DISCUSSION

It has been demonstrated that the techniques of computer vision and machine learning can predict the pulse rate, pulse rate variability, and force exertion level using extracted features and provides a novel approach for such estimation. Understanding force exertion levels has important implications across domains and applications, and in this work, the approach is demonstrated in the context of workplace injuries. Specifically, varying levels of force and duration/frequency of these forces are predictive of musculoskeletal injuries. This section provides more discussion on using machine learning in prediction of force exertion level and provides more insights on the selected features for the prediction of force exertion level.

6.1   Machine Learning in Classifying Force Levels

The use of machine learning is two-fold in this work. First, the machine learning is used in DeepFace algorithm for facial recognition that our team further augmented with increased number of features. Secondly, we use machine learning to generate a classifier to predict different force exertion levels.

There are various methodologies [66–71, 79, 80] proposed that can achieve facial recognition but the methodology proposed in [47] outperforms other methods and results in the accuracy of 97.35% in Labeled Faces in the Wild (LFW) dataset, reducing the error in face recognition of current state-of-the-art by more than 27%. The 9 layer neural network used in Deepface makes it more robust to detect faces in the video for our study and henceforth extract relevant features from the video frames. These facial features represent a key component for force classification.

The second use of machine learning is force classification. In this step, we added additional novelty by leveraging the underlying physiological mechanisms of generat-

ing muscle forces to improve force classification accuracy. Thus, features from PPG are included as well and deployed a fully connected neural network to train a model that can distinguish between different force exertion levels. The neural networks are known to be universal approximators [81] and hence we use them to identify the underlying function explaining the relationship between the features and response variable.

6.2 Facial Features Selection

The average movement of detected facial landmarks along with the cardiovascular features derived from PPG in different exertion levels have been used to classify the force exertion levels. The novelty lies in choosing these relevant facial features. As the person increases her effort level, facial expression tends to change and there are differences in the average movement of the facial landmarks for different force exertion level. The identification of these visual cues were drawn for tools and techniques from the field of human factors. Specifically, ergonomics practitioners are trained to associate (through observation) cues like "Substantial Effort with Changed Facial Expression" with an MVC of 70 % and very strong effort. In contrast, "Obvious Effort, But Unchanged Facial Expression" is associated with 40 % MVC and moderate effort [82]

Figure 6.2 & Figure 6.1 shows how different groups of facial landmarks behave differently for three different force exertion levels. Figure 6.1 shows the average movement of landmark groups for three randomly picked subjects. It is interesting to note that landmarks belonging to nose always shows least movement in all the three force exertion levels. On the other hand, face contour, eyes and cheeks show high average movements over the entire video. Figure 6.2 generalizes this behavior over all the subjects and depicts the box plot of each force exertion level for all the 7 groups of landmarks. The change in the location of the landmarks on the face is explained by the motion of the muscles beneath the skin of the face. As body changes its actions,

Fig. 6.1. Variation in facial features groups of three randomly chosen subjects, 1: Contour of a face, 2: Left Eye + Eyebrow, 3: Right Eye + Eyebrow, 4: Nose, 5: Lips,6: Left Cheek, 7: Right Cheek
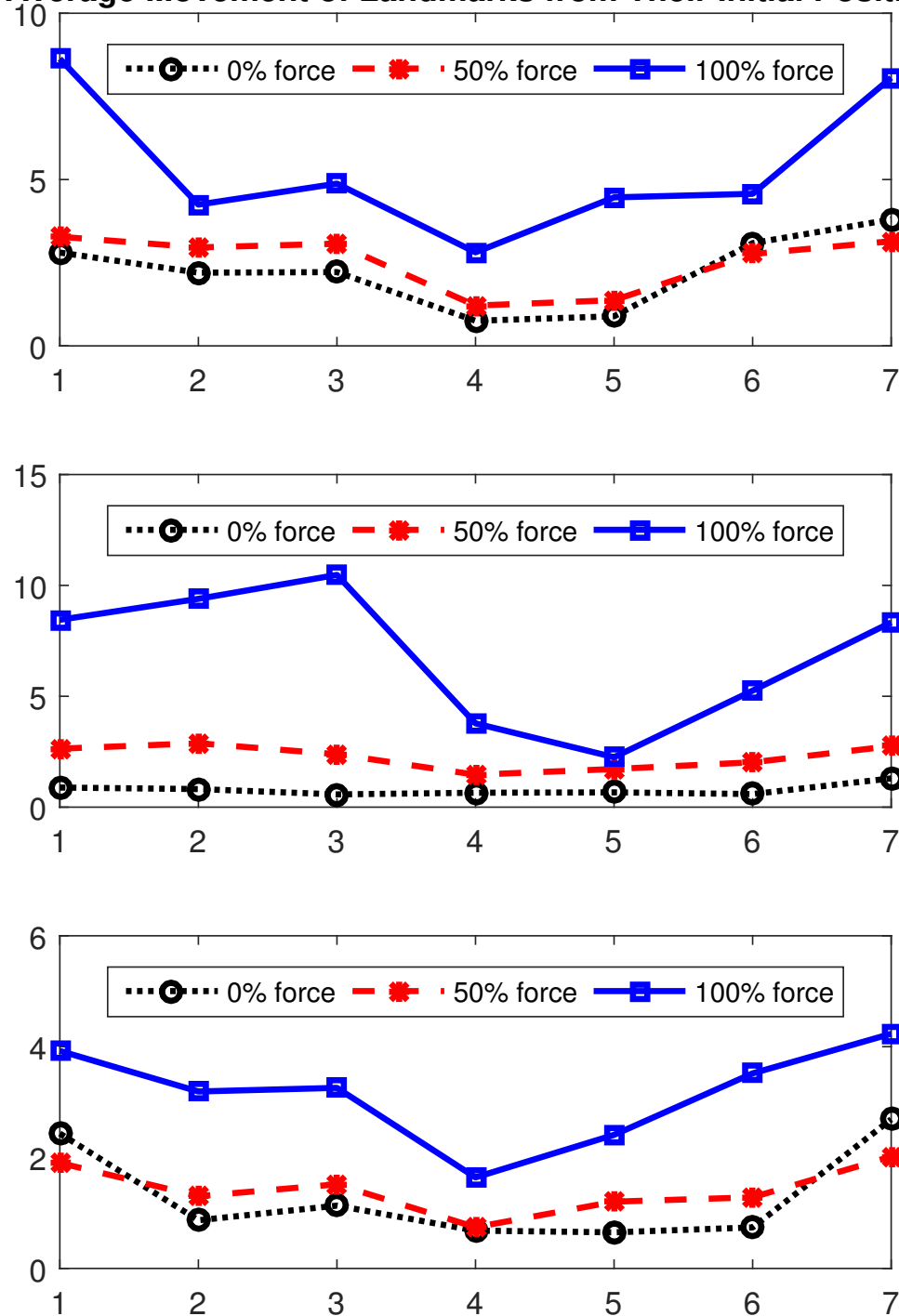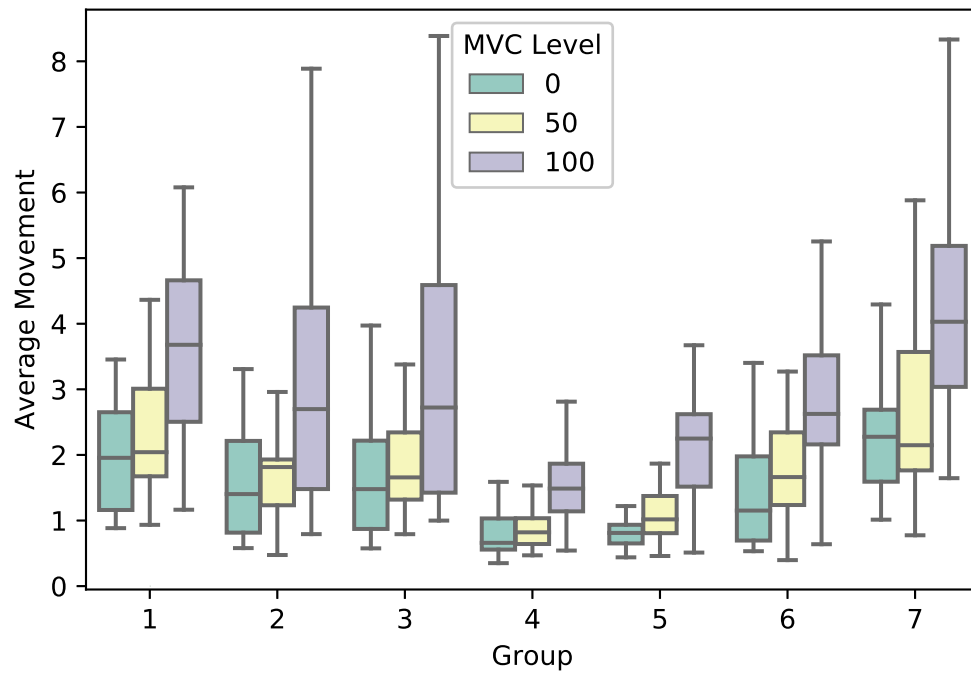
Fig. 6.2. Variation in facial features groups of all subjects, 1: Contour of a face, 2: Left Eye + Eyebrow, 3: Right Eye + Eyebrow, 4: Nose, 5: Lips,6: Left Cheek, 7: Right Cheek
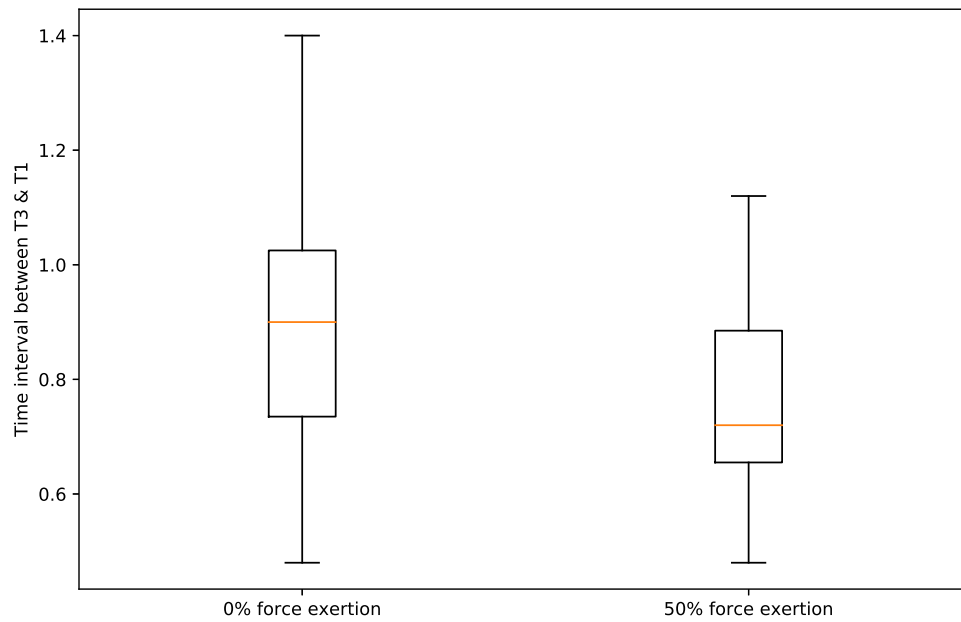
Fig. 6.3. Variation in the time interval between T3 & T1 for all subjects corresponding to 0% & 50% force exertion level

it leads to the changes in the facial expression [83] and thus we observe the movement of landmarks for different force exertion levels. It is further interesting to note that average movement of landmarks are robust against day to day variations like change in the lighting around them, presence of make-up on the face etc. as well as robust for different people belonging to different skin tone. Therefore, choice of such facial feature leads to high accuracy of our model and make it robust for classifying higher force exertion level.

As noted previously, changes in facial expressions are observed typically for strong exertions (60-70 % MVC); however, a known musculoskeletal injury mechanism is continuous and prolonged sub-maximal force exertions. Specifically, although a single moderate (30-50 % MVC) exertion may not lead to immediate injuries, repeated and prolonged exertions at these levels lead to cumulative trauma disorders. The key challenge is that facial expressions are more likely to be unchanged during these exertion levels and therefore average movement of landmarks is not used for classifying lower level of force exertion i.e., 50% and 0%. Thus, we further distinguished lower levels of force exertions using cardiovascular parameters of the person which would be captured from PPG signal. This changing trend can be seen in Figure 4.3 where the increasing trend of PPG signal is observed for 50% force exertion level and a stationary signal for 0% force exertion level. Figure 6.3 shows the observed variation in one of the PPG feature i.e., time interval between T3 and T1 for first four beats, for 0% and 50% force exertion levels. The mean and standard deviation of time interval between T3 and T1 for 0% effort level is 1.03s and 0.55s respectively where mean and standard deviation for 50% force exertion level is 0.86 sec and 0.52 sec respectively. Higher force exertion activity increases the heart rate of the person because of faster cardiac cycles, hence we see differences in PPG extracted features between 0% & 50% effort level. Therefore, both average movement and PPG features become important features in our study.

For our analysis, we utilized cardiovascular features derived from the PPG signal that had been captured using a contact device placed on the earlobe of the subjects.

Although this technique requires contact, continued innovation in wearables (e.g., fitness watches and activity trackers) has provided many options for collection of continuous PPG signals without significant cost to employers or usability/workflow burden to workers. Furthermore, over the last decade, there have been ongoing research in developing methods for estimating PPG signals from the facial videos using non contact methodology. The authors in [49] provided a technique that extracts PPG signal from the human facial videos using complimentary metal-oxide semiconductor camera with the use of external light emitting diodes. Also, the authors of [39, 40] demonstrated that PPG signal can be estimated by just using ambient light as a source of illumination along with simple digital camera. Further advancements led to the formulation of more robust methodology that overcomes challenges in extracting PPG for people having dark skin tones [42]. There are many existing methods that can be easily used to derive PPG signal directly from the facial videos. Future work incorporating these techniques have the potential to make this proposed methodology completely passive and non-contact.

6.3 Non-contact Exposure Assessment

The force exertions has been considered as one of the main contributing factors in current risk assessment tools [84–86]. The high variability of the identified risk score with respect to the estimated force exertion parameters is reported in current assessment tools. For example, the Strain Index Assessment [86] score will double if the intensity of the exertion changes from 20% to 40% [24]. In addition, [87] reported weak correlation values between the ergonomists estimates and the worker's self-reports for pinch and grip force. The proposed non-contact assessment method for classifying force levels can provide an objective automated estimations of hand forces.

6.4   Prediction of Pulse Rate

The use of machine learning in predicting pulse rate and pulse rate variability from the facial videos has been successfully demonstrated. The other researchers in the field have utilized videos of the person and estimate the PPG signal that was utilized to calculate the PR and PRV, but in this work, instead of predicting PPG signal, more granualar features of PPG: PR, & PRV have been predicted. In doing so, this work predicts the average value of PR over the whole video of 50 seconds. It is known that the PR of the person can change during the interval of 50 seconds, but this work doesn't predict the changing trend of the pulse rate over a period of time. However, the methodology discussed in this work may be used to predict pulse rate for shorter (such as 5 seconds) videos as well that can give us idea about how individuals pulse rate is changing over time. This work considered longer videos because predictability of PRV improves with the length of the video [64]. On the other hand, continuous prediction of the changing trend of pulse rate over a given time would be an interesting exercise and would require different approach and modeling scheme such as recurrent neural networks (RNNs) [88] for such predictions and is not in the scope of this work.

7. CONCLUSION

The monitoring of health parameters like PR, PRV, and exertion level is important to keep check on individual's health and spot the potential cardiovascular diseases and musculoskeletal injuries. Recently, the use of non-contact method such as using camera videos is preferred over contact methods like pulse oximeters and observational methods for such measurement. In this work, a computer vision and machine learning technique is leveraged to predict pulse rate, LF & HF component of pulse rate variability, and force exertion level. The physiological parameters are remotely predicted using the video of human face captured using a laptop's camera. The subtle changes in the face pixels intensity over the different frames of the video are exploited to train a neural network with three hidden layers. Because computer vision is not intrusive to the workers and can be done without the need for specialize equipment, this technique will provide workplaces a transformative tool for ensuring on-the-job cardiovascular problems and force requirements (effort level and duration at these levels) do not contribute to workplace injuries. Experimental evaluations are performed for twenty subjects, and the proposed approach demonstrates significant improvement as compared to the baselines thus validating that the approach has the potential to be applied in real scenarios. Although current work accurately classifies three force exertion levels, work is ongoing to expand this technique to other exertion levels to better meet the varying needs of different workplaces.

REFERENCES

REFERENCES

[1] A. I. Aladin, S. P. Whelton, M. H. Al-Mallah, M. J. Blaha, S. J. Keteyian, S. P. Juraschek, J. Rubin, C. A. Brawner, and E. D. Michos, "Relation of resting heart rate to risk for all-cause mortality by gender after considering exercise capacity (the henry ford exercise testing project)," *The American journal of cardiology*, vol. 114, no. 11, pp. 1701–1706, 2014.

[2] Å. Hjalmarson, E. A. Gilpin, J. Kjekshus, G. Schieman, P. Nicod, H. Henning, and J. Ross, "Influence of heart rate on mortality after acute myocardial infarction," *The American journal of cardiology*, vol. 65, no. 9, pp. 547–553, 1990.

[3] D. H. Spodick, "Survey of selected cardiologists for an operational definition of normal sinus heart rate," *The American journal of cardiology*, vol. 72, no. 5, pp. 487–488, 1993.

[4] J. W. Mason, D. J. Ramseth, D. O. Chanter, T. E. Moon, D. B. Goodman, and B. Mendzelevski, "Electrocardiographic reference ranges derived from 79,743 ambulatory subjects," *Journal of electrocardiology*, vol. 40, no. 3, pp. 228–234, 2007.

[5] K. Fox, I. Ford, P. G. Steg, M. Tendera, M. Robertson, R. Ferrari *et al.*, "Heart rate as a prognostic risk factor in patients with coronary artery disease and left-ventricular systolic dysfunction (beautiful): a subgroup analysis of a randomised controlled trial," *The Lancet*, vol. 372, no. 9641, pp. 817–821, 2008.

[6] G. G. Berntson, J. Thomas Bigger, D. L. Eckberg, P. Grossman, P. G. Kaufmann, M. Malik, H. N. Nagaraja, S. W. Porges, J. P. Saul, P. H. Stone *et al.*, "Heart rate variability: origins, methods, and interpretive caveats," *Psychophysiology*, vol. 34, no. 6, pp. 623–648, 1997.

[7] P. Nickel and F. Nachreiner, "Sensitivity and diagnosticity of the 0.1-hz component of heart rate variability as an indicator of mental workload," *Human Factors*, vol. 45, no. 4, pp. 575–590, 2003.

[8] P. Jönsson, "Respiratory sinus arrhythmia as a function of state anxiety in healthy individuals," *International journal of psychophysiology*, vol. 63, no. 1, pp. 48–54, 2007.

[9] N. Hjortskov, D. Rissén, A. K. Blangsted, N. Fallentin, U. Lundberg, and K. Søgaard, "The effect of mental stress on heart rate variability and blood pressure during computer work," *European journal of applied physiology*, vol. 92, no. 1-2, pp. 84–89, 2004.

[10] S. Akselrod, D. Gordon, F. A. Ubel, D. C. Shannon, A. C. Barger, and R. J. Cohen, "Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control," *science*, pp. 220–222, 1981.

[11] U. B. of Labor Statistics, "Survey of occupational injuries and illnesses, in coop-
    eration with participating state agencies," *U.S Department of Labor*, 2016.

[12] R. Wells, "Why have we not solved the msd problem?" *Work*, vol. 34, no. 1, pp.
    117–121, 2009.

[13] M. J. Hoozemans, A. J. Van Der Beek, M. H. Fringsdresen, F. J. Van Dijk,
    and L. H. Van Der Woude, "Pushing and pulling in relation to musculoskeletal
    disorders: a review of risk factors," *Ergonomics*, vol. 41, no. 6, pp. 757–781, 1998.

[14] B. R. da Costa and E. R. Vieira, "Risk factors for work-related musculoskeletal
    disorders: a systematic review of recent longitudinal studies," *American journal
    of industrial medicine*, vol. 53, no. 3, pp. 285–323, 2010.

[15] P. W. Buckle and J. J. Devereux, "The nature of work-related neck and upper
    limb musculoskeletal disorders," *Applied ergonomics*, vol. 33, no. 3, pp. 207–217,
    2002.

[16] W. M. Keyserling, "Workplace risk factors and occupational musculoskeletal
    disorders, part 2: A review of biomechanical and psychophysical research on risk
    factors associated with upper extremity disorders," *AIHAJ-American Industrial
    Hygiene Association*, vol. 61, no. 2, pp. 231–243, 2000.

[17] S. P. Schneider, "Musculoskeletal injuries in construction: a review of the lit-
    erature," *Applied occupational and environmental hygiene*, vol. 16, no. 11, pp.
    1056–1064, 2001.

[18] K. G. Hauret, B. H. Jones, S. H. Bullock, M. Canham-Chervak *et al.*, "Mus-
    culoskeletal injuries: description of an under-recognized injury problem among
    military personnel," *American journal of preventive medicine*, vol. 38, no. 1, pp.
    S61–S70, 2010.

[19] E. Koppelaar and R. Wells, "Comparison of measurement methods for quanti-
    fying hand force," *Ergonomics*, vol. 48, no. 8, pp. 983–1007, 2005.

[20] P. W. Buckle and J. Jason Devereux, "The nature of work-related neck and
    upper limb musculoskeletal disorders," *Applied Ergonomics*, vol. 33, no. 3,
    pp. 207–217, 2002. [Online]. Available: https://www.sciencedirect.com/science/
    article/pii/S0003687002000145

[21] W. M. Keyserling, "Workplace Risk Factors and Occupational Musculoskeletal
    Disorders, Part 2: A Review of Biomechanical and Psychophysical Research on
    Risk Factors Associated with Upper Extremity Disorders," *AIHAJ - American
    Industrial Hygiene Association*, vol. 61, no. 2, pp. 231–243, mar 2000. [Online].
    Available: http://www.tandfonline.com/doi/abs/10.1080/15298660008984532

[22] S. P. Schneider, "Musculoskeletal injuries in construction: A review
    of the literature," pp. 1056–1064, nov 2001. [Online]. Available: http:
    //www.tandfonline.com/doi/abs/10.1080/104732201753214161

[23] K. G. Hauret, B. H. Jones, S. H. Bullock, M. Canham-Chervak, and
    S. Canada, "Musculoskeletal injuries: Description of an under-recognized
    injury problem among military personnel," *American Journal of Preventive
    Medicine*, vol. 38, no. 1 SUPPL., pp. S61–S70, jan 2010. [Online]. Available:
    https://www.sciencedirect.com/science/article/pii/S0749379709006746

[24] E. Koppelaar and R. Wells, "Comparison of measurement methods for quantifying hand force," *Ergonomics*, vol. 48, no. 8, pp. 983–1007, jun 2005. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00140130500120841

[25] National Institute for Occupational Safety and Health (NIOSH), "Musculoskeletal disorders and workplace factors: a critical review of epidemiologic evidence for WMSDs of the neck, upper extremity and low back," 1997. [Online]. Available: https://stacks.cdc.gov/view/cdc/21745

[26] A. Schwarzer, "The prevalence and clinical features of internal disc disruption in patients with chronic lbp," *Spine*, vol. 20, pp. 1878–1883, 1995.

[27] M. A. Adams, B. J. Freeman, H. P. Morrison, I. W. Nelson, and P. Dolan, "Mechanical initiation of intervertebral disc degeneration," *Spine*, vol. 25, no. 13, pp. 1625–1636, 2000.

[28] B. Fung, K. Chan, L. Lam, S. Cheung, N. Choy, K. Chu, L. Chung, W. Liu, K. Tai, S. Yung *et al.*, "Study of wrist posture, loading and repetitive motion as risk factors for developing carpal tunnel syndrome," *Hand surgery*, vol. 12, no. 01, pp. 13–18, 2007.

[29] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological measurement*, vol. 28, no. 3, p. R1, 2007.

[30] J. M. Haynes, "The ear as an alternative site for a pulse oximeter finger clip sensor," *Respiratory care*, vol. 52, no. 6, pp. 727–729, 2007.

[31] G. Borg, "Psychophysical scaling with applications in physical work and the perception of exertion," *Scandinavian journal of work, environment & health*, pp. 55–58, 1990.

[32] D. S. Stetson, B. A. Silverstein, W. M. Keyserling, R. A. Wolfe, and J. W. Albers, "Median sensory distal amplitude and latency: comparisons between nonexposed managerial/professional employees and industrial workers," *American Journal of Industrial Medicine*, vol. 24, no. 2, pp. 175–189, 1993.

[33] J. S. Casey, R. W. McGorry, and P. G. Dempsey, "Getting a grip on grip force estimates: A valuable tool for ergonomic evaluations," *Professional Safety*, vol. 47, no. 10, p. 18, 2002.

[34] R. W. Bohannon, A. Peolsson, N. Massy-Westropp, J. Desrosiers, and J. Bear-Lehman, "Reference values for adult grip strength measured with a jamar dynamometer: a descriptive meta-analysis," *Physiotherapy*, vol. 92, no. 1, pp. 11–15, 2006.

[35] P. J. Keir and J. P. Mogk, "The development and validation of equations to predict grip force in the workplace: contributions of muscle activity and posture," *Ergonomics*, vol. 48, no. 10, pp. 1243–1259, 2005.

[36] S. N. Sidek and A. J. H. Mohideen, "Mapping of emg signal to hand grip force at varying wrist angles," in *Biomedical Engineering and Sciences (IECBES), 2012 IEEE EMBS Conference on*. IEEE, 2012, pp. 648–653.

[37] Z. J. Fan, B. A. Silverstein, S. Bao, D. K. Bonauto, N. L. Howard, and C. K. Smith, "The association between combination of hand force and forearm posture and incidence of lateral epicondylitis in a working population," *Human factors*, vol. 56, no. 1, pp. 151–165, 2014.

[38] P. Spielholz, B. Silverstein, M. Morgan, H. Checkoway, and J. Kaufman, "Comparison of self-report, video observation and direct measurement methods for upper extremity musculoskeletal disorder physical risk factors," *Ergonomics*, vol. 44, no. 6, pp. 588–613, 2001.

[39] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light." *Optics express*, vol. 16, no. 26, pp. 21434–21445, 2008.

[40] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE transactions on biomedical engineering*, vol. 58, no. 1, pp. 7–11, 2011.

[41] Y. Sun, S. Hu, V. Azorin-Peris, S. Greenwald, J. Chambers, and Y. Zhu, "Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise," *Journal of biomedical optics*, vol. 16, no. 7, pp. 077010–077010, 2011.

[42] M. Kumar, A. Veeraraghavan, and A. Sabharwal, "Distanceppg: Robust noncontact vital signs monitoring using a camera," *Biomedical optics express*, vol. 6, no. 5, pp. 1565–1588, 2015.

[43] M. A. Haque, R. Irani, K. Nasrollahi, and T. B. Moeslund, "Heartbeat rate measurement from facial video," *IEEE Intelligent Systems*, vol. 31, no. 3, pp. 40–48, May 2016.

[44] X. Li, J. Chen, G. Zhao, and M. Pietikinen, "Remote heart rate measurement from face videos under realistic situations," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 4264–4271.

[45] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rppg," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, Oct 2013.

[46] X. Yu, J. Huang, S. Zhang, and D. N. Metaxas, "Face landmark fitting via optimized part mixtures and cascaded deformable model," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 11, pp. 2212–2226, 2016.

[47] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1701–1708.

[48] F. P. Wieringa, F. Mastik, and A. F. W. v. d. Steen, "Contactless multiple wavelength photoplethysmographic imaging: A first step toward "spo2 camera" technology," *Annals of Biomedical Engineering*, vol. 33, no. 8, pp. 1034–1041, Aug 2005. [Online]. Available: https://doi.org/10.1007/s10439-005-5763-2

[49] K. Humphreys, T. Ward, and C. Markham, "Noncontact simultaneous dual wavelength photoplethysmography: a further step toward noncontact pulse oximetry," *Review of scientific instruments*, vol. 78, no. 4, p. 044304, 2007.

[50] B. D. Holton, K. Mannapperuma, P. J. Lesniewski, and J. C. Thomas, "Signal recovery in imaging photoplethysmography," *Physiological measurement*, vol. 34, no. 11, p. 1499, 2013.

[51] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation." *Optics express*, vol. 18, no. 10, pp. 10 762–10 774, 2010.

[52] A. Lam and Y. Kuno, "Robust heart rate measurement from video using select random patches," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015, pp. 3640–3648.

[53] A. Osman, J. Turcot, and R. El Kaliouby, "Supervised learning approach to remote heart rate estimation from facial videos," in *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, vol. 1. IEEE, 2015, pp. 1–6.

[54] G.-S. Hsu, A. Ambikapathi, and M.-S. Chen, "Deep learning with time-frequency representation for pulse estimation from facial videos," in *Biometrics (IJCB), 2017 IEEE International Joint Conference on*. IEEE, 2017, pp. 383–389.

[55] C. H. Chen, Y. H. Hu, T. Y. Yen, and R. G. Radwin, "Automated video exposure assessment of repetitive hand activity level for a load transfer task," *Human Factors*, vol. 55, no. 2, pp. 298–308, apr 2013. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0018720812458121

[56] R. L. Greene, D. P. Azari, Y. H. Hu, and R. G. Radwin, "Visualizing stressful aspects of repetitive motion tasks and opportunities for ergonomic improvements using computer vision," *Applied Ergonomics*, vol. 65, pp. 461–472, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S000368701730056X

[57] O. Akkas, C.-H. Lee, Y. H. Hu, T. Y. Yen, and R. G. Radwin, "Measuring elemental time and duty cycle using automated video processing," *Ergonomics*, vol. 59, no. 11, pp. 1514–1525, nov 2016. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/00140139.2016.1146347

[58] M. Liu, S. Han, and S. Lee, "Tracking-based 3d human skeleton extraction from stereo video camera toward an on-site safety and ergonomic analysis," *Construction Innovation*, vol. 16, no. 3, pp. 348–367, 2016.

[59] J. Seo, R. Starbuck, S. Han, S. Lee, and T. J. Armstrong, "Motion data-driven biomechanical analysis during construction tasks on sites," *Journal of Computing in Civil Engineering*, vol. 29, no. 4, p. B4014005, 2014.

[60] R. Starbuck, J. Seo, S. Han, and S. Lee, "A stereo vision-based approach to marker-less motion capture for on-site kinematic modeling of construction worker tasks," in *Computing in Civil and Building Engineering (2014)*, 2014, pp. 1094–1101.

[61] M. Elgendi, "On the analysis of fingertip photoplethysmogram signals," *Current cardiology reviews*, vol. 8, no. 1, pp. 14–25, 2012.

[62] J. Dorlas and J. Nijboer, "Photo-electric plethysmography as a monitoring device in anaesthesia: application and interpretation," *BJA: British Journal of Anaesthesia*, vol. 57, no. 5, pp. 524–530, 1985.

[63] N. A. Shirwany and M.-h. Zou, "Arterial stiffness: a brief review," *Acta Pharmacologica Sinica*, vol. 31, no. 10, p. 1267, 2010.

[64] W.-H. Lin, D. Wu, C. Li, H. Zhang, and Y.-T. Zhang, "Comparison of heart rate variability from ppg with that from ecg," in *The International Conference on Health Informatics*. Springer, 2014, pp. 213–215.

[65] B. WELLNESS, *Your Heart Beat and Your Health*, 2010 (accessed November 3, 2018), http://www.berkeleywellness.com/fitness/exercise/article/your-heart-beat-and-your-health.

[66] O. Barkan, J. Weill, L. Wolf, and H. Aronowitz, "Fast high dimensional vector multiplication face recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1960–1967.

[67] X. Cao, D. Wipf, F. Wen, G. Duan, and J. Sun, "A practical transfer learning algorithm for face verification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 3208–3215.

[68] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3025–3032.

[69] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 539–546.

[70] G. B. Huang, H. Lee, and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2518–2525.

[71] Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3476–3483.

[72] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, 2013, vol. 112.

[73] M. L. Bermingham, R. Pong-Wong, A. Spiliopoulou, C. Hayward, I. Rudan, H. Campbell, A. F. Wright, J. F. Wilson, F. Agakov, P. Navarro *et al.*, "Application of high-dimensional feature selection: evaluation for genomic prediction in man," *Scientific reports*, vol. 5, p. 10312, 2015.

[74] D. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," *CoRR*, vol. abs/1511.07289, 2015. [Online]. Available: http://arxiv.org/abs/1511.07289

[75] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, 2015. [Online]. Available: http://arxiv.org/abs/1502.03167

[76] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[77] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980

[78] R. Caruana, S. Lawrence, and C. L. Giles, "Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping," in *Advances in neural information processing systems*, 2001, pp. 402–408.

[79] P. J. Phillips, J. R. Beveridge, B. A. Draper, G. Givens, A. J. O'Toole, D. S. Bolme, J. Dunlop, Y. M. Lui, H. Sahibzada, and S. Weimer, "An introduction to the good, the bad, amp; the ugly face recognition challenge problem," in *Face and Gesture 2011*, March 2011, pp. 346–353.

[80] M. Osadchy, "Synergistic face detection and pose estimation with energy-based models," *Journal of machine learning research : JMLR.*, 2007.

[81] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, vol. 2, no. 5, pp. 359–366, 1989.

[82] T. J. Armstrong, "Acgih tlv for hand activity level," in *Biomechanics in Ergonomics*. CRC Press, 2007, pp. 377–390.

[83] C. Fantoni and W. Gerbino, "Body actions change the appearance of facial expressions," *PLoS One*, vol. 9, no. 9, 2014. [Online]. Available: http://search.proquest.com/docview/1564617673/

[84] S. Hignett and L. McAtamney, "Rapid entire body assessment," in *Handbook of Human Factors and Ergonomics Methods*. CRC Press, 2004, pp. 97–108.

[85] L. McAtamney and E. N. Corlett, "Rula: a survey method for the investigation of work-related upper limb disorders," *Applied ergonomics*, vol. 24, no. 2, pp. 91–99, 1993.

[86] J. Steven Moore and A. Garg, "The strain index: a proposed method to analyze jobs for risk of distal upper extremity disorders," *American Industrial Hygiene Association Journal*, vol. 56, no. 5, pp. 443–458, 1995.

[87] S. Bao, N. Howard, P. Spielholz, and B. Silverstein, "Quantifying repetitive hand activity for epidemiological research on musculoskeletal disorders–part ii: comparison of different methods of measuring force level and repetitiveness," *Ergonomics*, vol. 49, no. 4, pp. 381–392, 2006.

[88] S. Grossberg, "Recurrent neural networks," *Scholarpedia*, vol. 8, no. 2, p. 1888, 2013.