VIDEO BASED ATTENTION ASSESSMENT

Authors

Nishant Mishra: 2018165

Himanshu Kumar: 2018147

BTP report submitted in partial fulfillment of the requirements
for the Degree of B.Tech. in Electronics & Communication Engineering (Nishant Mishra) and
Electronics & Communication Engineering (Himanshu Kumar)

on May 12, 2021

BTP Track: Research Track

BTP Advisor

Dr. Sujay Deb

Indraprastha Institute of Information Technology

New Delhi

Student's Declaration

We hereby declare that the work presented in the report entitled **"Video-Based Attention Assessment"** submitted by us for the partial fulfillment of the requirements for the degree of *Bachelor of Technology* in *Electronics & Communication Engineering* at Indraprastha Institute of Information Technology, Delhi, is an authentic record of our work carried out under the guidance of **Dr. Sujay Deb**. Due acknowledgments have been given in the report to all material used. This work has not been submitted anywhere else for the reward of any other degree.

..... **Place & Date: .....**
Nishant Mishra

..... **Place & Date: .....**
Himanshu Kumar

Certificate

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

..... **Place & Date: .....**
Dr. Sujay Deb

Abstract

Attention Assessment is one of the most interesting topic of research in the ongoing era. Many researches going on in this field to develop a tool that can accurately measure the attention level. The past researches on this topic had shown a strong relationship between heart rate variability (HRV), heart rate (HR), and cognitive state of mind (Stress/Attention). Researchers have used traditional heart rate calculation methods like using ECG signal and photoplethysmography (PPG) signal in the previous approaches. However, the traditional approaches of finding heart rate are not user-friendly, as the calculation of ECG signal involves a complex procedure; at the same time, these methods are not cost-friendly. In recent time many researchers are trying to measure heart rate using video-based remote PPG extraction methods; these methods do not require any contact and can be measured remotely so that they can be beneficial in fields like telemedicine. In the past researcher has tried different methods to find heart rate using the video-based approach, some have used a deep learning-based approach, and others have used signal processing-based approaches. At the same time, there is no significant research on finding the attention level from scratch using a video-based approach is there till now. This research aims to excess the strengths and weaknesses of these existing models and theories and tries to build a video-based attention assessment system using computer vision and signal processing, which will be reliable and at the same time cost-effective. We also intend to make a platform(Android App/Web extension) that will increase the accessibility of our product in day-to-day life.

Keywords: Attention Assessment,Heart Rate (HR), Heart Rate Variability (HRV), Computer Vision, Signal Processing

Acknowledgments

We want to express our special thanks of gratitude to Dr. Sujay. Deb, for giving us the opportunity to work on this wonderful project. We also thank him for his invaluable suggestions on how to approach the research problem.

We are also grateful to the Indraprastha Institute of Information Technology, Delhi, for providing us with the resources and equipment for doing this project.

Work Distribution

The work was divided equally among us, and we have approached every problem together so to solve issues efficiently. To be specific, Himanshu Kumar played the main role in sections 3.2.2, 3.2.3, and 3.2.4. At the same time, Nishant Mishra played the main role in section 3.2.5, and both of us have equally contributed in report making and model testing part. The research required to proceed with the project was done jointly by both the members.

Contents

Chapter 1

Introduction

In the 21st century, where the population is skyrocketed and everybody faces huge pressure and stress to survive and make a place in their field of expertise, the true attention in work will be a deciding factor. At the same time, there are many professionals, as defense and security personnel, pilots, air traffic controllers, etc., where a high level of attention is necessary to develop their work in safe conditions. However, this attention may be affected, among others, by psychological stress and by states of sleep deprivation. According to research done on the field of finding a correlation between attention and performance reveals that nearly 60% of Individuals are not able to work with their 100% strength because they are not able to maintain the attention in their work which leads to a decline in productivity and hence affect their professional and personal life.

There are many previous research work like in[9], [2] and [3] which shows the relationship between heart rate variability (a function of heart rate), heart rate, and cognitive state of mind, the fundamental hypothesis on which this work is based consists in the fact that alterations in the autonomic nervous system (ANS) during the execution of an activity that requires the subject sustained attention can be noninvasively quantified by the recording of physiological signals. These alterations in the ANS can be studied by analyzing the heart rate variability (HRV) from the electrocardiographic (ECG) signal or the Pulse Rate Variability (PRV) from the photoplethysmographic (PPG) signal. The ANS is composed of two branches, the sympathetic nervous system, and the parasympathetic or vagal nervous system. HRV or PRV spectral analysis reveals two main components: a high-frequency (HF) component due to respiratory sinus arrhythmia, and a low-frequency (LF) component, which reflects both sympathetic and parasympathetic activity [9]. Power in the HF band has been used as a measure of parasympathetic activity. Normalized power in the LF band and the ratio between power in LF and HF bands have been considered as a measure of sympathovagal balance, so in summary, we can use the LF and HF components of the HRV signal to find the attention level of the subject.

In this research, we will use the rPPG method to extract the heart rate variability, and then we will use signal processing methods to find high frequency(HF) and low frequency(LF) components of it. Once we have these components, we will use the established research works to predict the level of attention.

1.1 Problem Statement

The major goal of this research work is to develop a cost-effective and easy to use remote video-based attention assessment system which further can be integrated on web or android application to track the attention level and provide a continuous analysis report.

1.2 Motivation

According to the latest Happiness Index made by the united nations(UN), India ranked 144 out of 156 countries, and this is going bad year after year (Figure 1.1), which shows that our people are not happy and live in a stressed situation which leads to low attention level in their work. In the current scenario, covid has worked as a catalyst to make the situation worse. The low attention level decreases the productivity of an individual and which leads to low production. In the covid era where all work like schools, colleges, and offices is working from home, and people-to-people interaction has been reduced, there is no tool by which a mentor or a friend can access the attention level and provide the necessary help and guidance to excel their work. If we have a reliable attention assessment system that can work remotely, then we can track the attention level of the individual remotely and use that to guide them, which can increase productivity and make them happier.

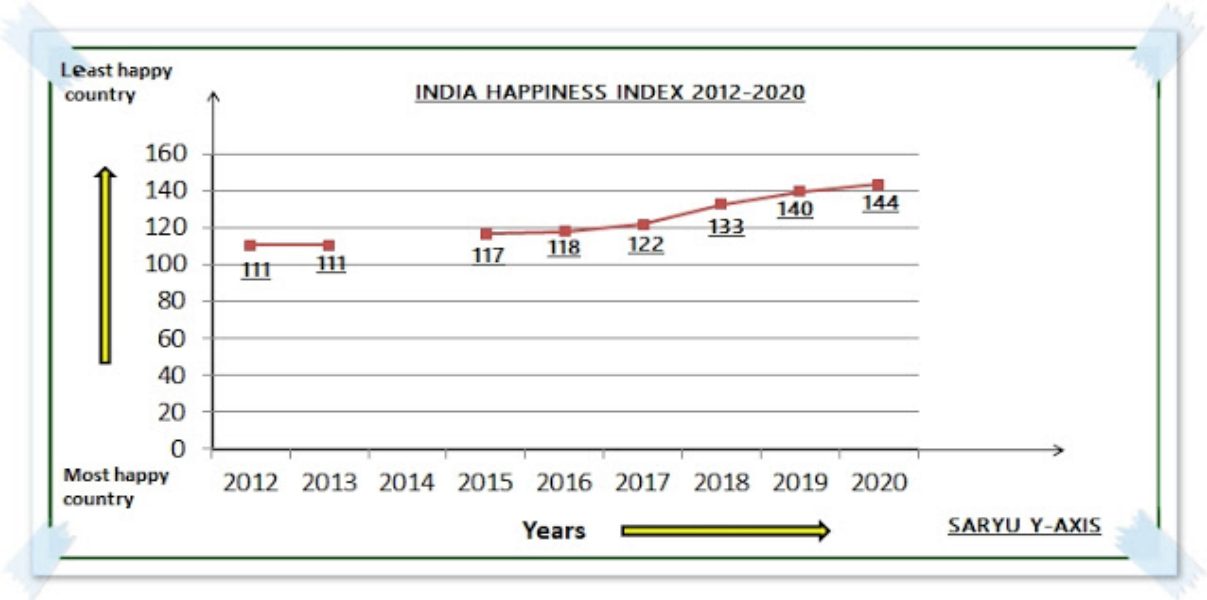

Figure 1.1: Happiness index Of India over the year(2012-20)

-

Chapter 2

Literature Review

In this section, we will discuss the papers we have read and analyzed to start our work. This section is divided into two parts. In section 2.1, we will discuss the relationship between attention assessment and heart rate variability (HRV), and in section 2.2, we will focus on getting these parameters from a video stream.

2.1 Relation Between Heart Rate Variability and Attention

Researchers have suggested that identifying the quality of learning by monitoring the cognitive states of the students is a task of value. Using these, teachers can adjust their teaching plans and methods in real-time from the student cognitive response to improve the students learning.

Traditionally, teachers rely on their experience, which may not be reliable in certain cases. Another important factor is that as we are moving towards an online teaching setting in which teachers cannot analyze many video streams of students for assessing their attention level during the online classes. Now, we will focus on a solution from which we develop an automatic attention assessment system.

There is a common understanding of what Attention is, according to [1], Attention is a selection of salient information and the allocation of cognitive processing appropriate to that information. It is known that bioelectric activities reflect behavioral and cognitive processes. We have evidence that the autonomic nervous system (ANS) correlates with the change in attention. So, the fundamental hypothesis on which this work is based consists in the fact that alterations in the Autonomic Nervous System (ANS) during the execution of an activity that requires the subject sustained attention can be noninvasively quantified by the recording of physiological signals. These alterations in the ANS can be studied by analyzing the heart rate variability (HRV) from the electrocardiographic (ECG) signal or the pulse rate variability (PRV) from the photoplethysmographic (PPG) signal. The ANS is composed of two branches, the sympathetic nervous system, and the parasympathetic or vagal nervous system. HRV or PRV spectral analysis reveals two main components: a high-frequency (HF) component due to respiratory

sinus arrhythmia, and a low-frequency (LF) component, which reflects both sympathetic and parasympathetic activity. Power in the HF band has been used as a measure of parasympathetic activity. Normalized power in the LF band and the ratio between power in LF and HF bands have been considered as a measure of sympathovagal balance.

Above was the theoretical basis of our work. We have also found experimental proof for the same[8]; in this work, they have performed an experiment with 11 healthy subjects, and they have performed three visual tasks during which their blood hemoglobin level data in the brain, respiratory data, and ECG data were collected to analyze the relationship between workload and bioelectric signal. Following are the tasks that subjects have to perform.

1. N Back Task: The n-back task is a representative task used to evaluate working memory. Subjects were shown a screen with a black or red number in the middle and were asked to provide the number presented in trials previously if the currently presented number was red. In order to manipulate the workload, participants performed 1-back and 2-back trials (low and high workload, respectively).

2. Stroop Task: In the Stroop task, subjects were required to discriminate three different meanings (Figure 2.1). Subjects were required to provide the number of groups in which the color, the object, and the word all matched (Figure 2.1 , left). In order to manipulate the workload, either two or four groups were displayed on the screen (low and high workload, respectively).

3. Visual search task: In the visual search task, subjects were required to find a target character on the screen and push a button when they find the target character. To manipulate the workload, the number of distractor inhibitions was changed: more distractor inhibitions created a low-load visual search, and fewer distractor inhibitions created a high-load visual search (low and high workload, respectively).

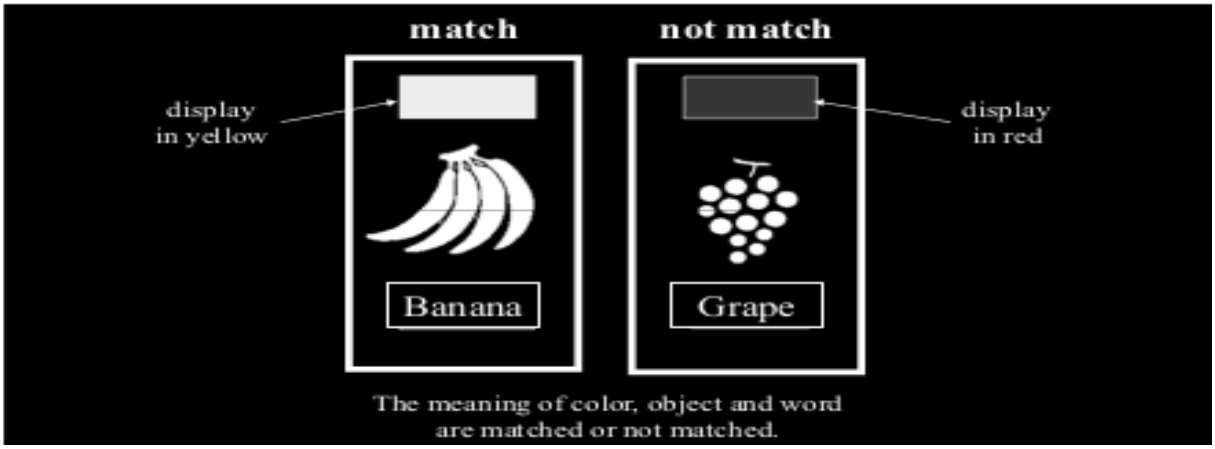

Figure 2.1: Representation of Stroop test

The results of the experiment show that HRV indices showed significant trends in the high workload. In high workload (means high cognitive and attentional state) condition, there was a decrease in the HF component of ECG signal and in mean RR interval. Therefore, the paper [8] has proposed that psychological stress caused by overload affects the HRV indices largely.

Now that we know theoretically and experimentally, there is a correlation between HRV and the attention level of an individual. We will focus on understanding and calculating these bioelectric signals. We will be using the PPG signal as our main signal to assess attention as PPG recording is very convenient. In contrast, the recording of the ECG signal involves several electrodes all over the chest of the subject. In the next sections, we will discuss different ways of recording PPG signals.

2.2 Heart Rate (HR) and Heart Rate Variability (HRV) Measurement

Traditionally heart rate is calculated using an ECG signal using of which is a very complex process. In recent time there is a trend of using heart rate sensors which is easy to use than earlier and give quite good results, but the issue with this method is it involves the use of sensors. Now the sensors which give good results (like sensors in Apple and Fitbit watches) are costly, and hence its accessibility is quite low. There is another alternative which is using PPG signal to find out the heart rate and hence heart rate variability.

2.2.1 Relation Between ECG and PPG Signal

Photoplethysmography (PPG) is a vascular optical measurement technique used to detect blood volume changes in the microvascular bed of target tissue. The finger and toe pad sites are usually assessed. A range of features of the pulse wave has been studied, including pulse transit time, pulse interval, peak-to-peak interval, amplitude, pulse contour, as well as their natural variability.

The (figure 2.2) shows the relationship between ECG and PPG signal, and It shows that there is a correlation between the peak of the PPG signal and the peak of ECG signal. Now, as the ECG signal gives real-time peak data but it's not the case with PPG, and it gives peaks after some delay, but the delay is fixed. This figure also confirms that the PPG signal will also be a good indicator of different features like RR (peak in case of PPG is called R, so its peak to peak) interval and HF (High Frequency) and LF (Low Frequency) components. The Paper [5] also confirms that there is a very close relation between PPG signal and ECG signal, and PPG can be used as an alternative to ECG.

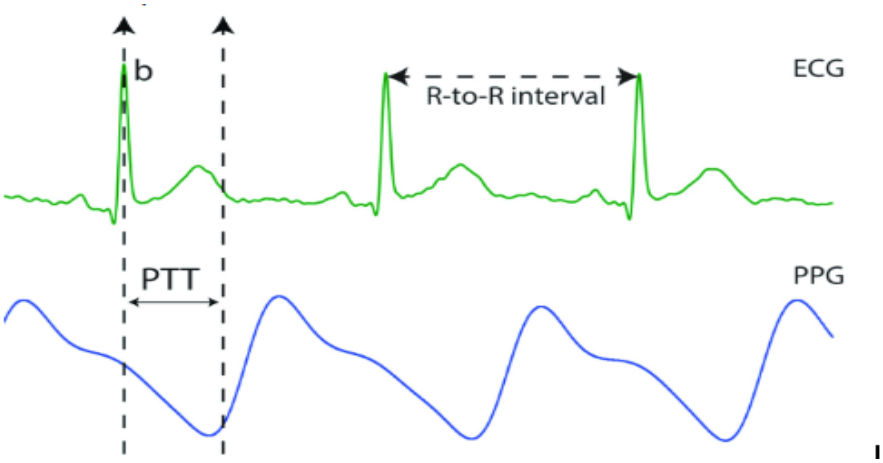

Figure 2.2: PPG signal and its correlation with ECG signal

2.2.2 PPG Signal Extraction Methods

As It is now established that the PPG signal can also be used to estimates the heart rate. So in this section, we will talk about its extraction method.

1. **Using Pulse Oximeter** one of the best methods we have found to extract PPG signal is using Spo2 Oximeter; a pictorial view on working of oximeter is shown in the figure below.

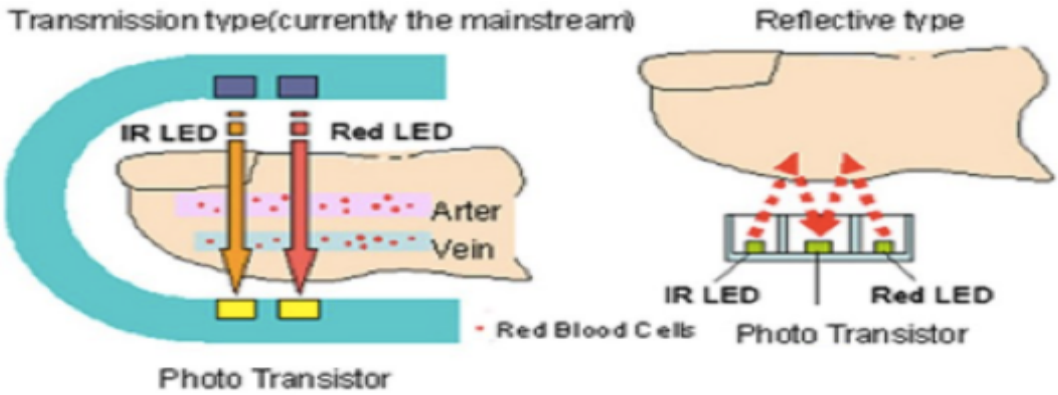

Figure 2.3: PPG signal Extraction using Oximeter

2. **Using Webcam and video processing techniques** is the most cost-effective method of all of the methods which we have discussed till now. This method only requires a device with a decent camera, and we can easily extract the required PPG signal. We have done an in-depth literature survey on the extraction of rPPG signal, and we will discuss it in the next section.

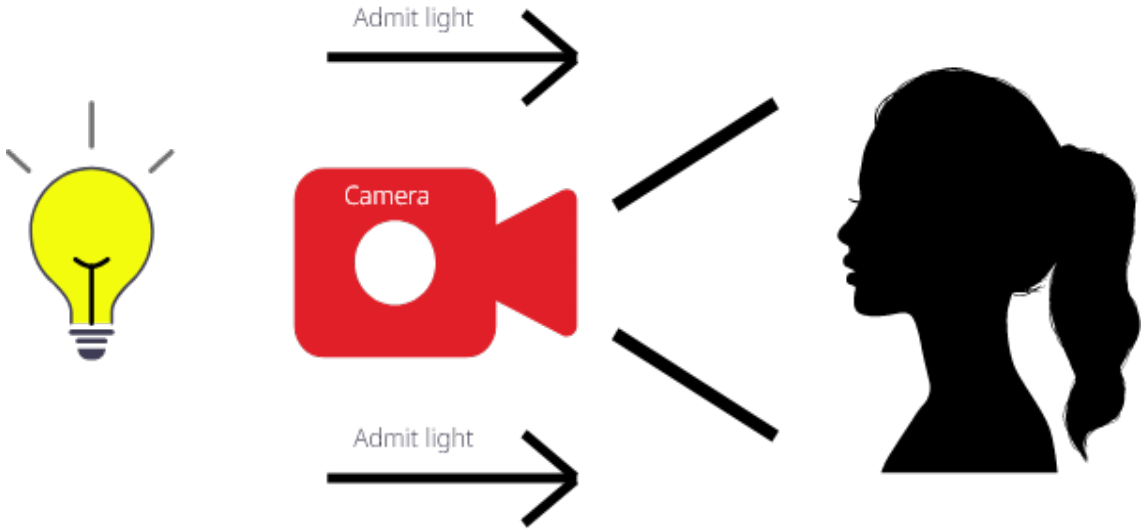

Figure 2.4: rPPG signal Extraction using Camera

2.2.3 rPPG Signal Extraction

In this subsection, we will cover all the theory which we have studies during literature survey about rPPG signal extraction using a webcam and the related model involved in it.

As we know, a camera captures an image of the objects through the light, which illuminates it and travels back to the camera from a light source after reflection from the object. In our case, the object is our skin, so we will try to model this skin reflection and analyze the underlying concept of extracting PPG remotely.

Skin Reflection Model We have found a detailed analysis of the Skin Reflection Model in the paper [10]. To model, a skin reflection phenomenon considers a light source illuminating a piece of human skin tissue and a remote color camera recording this image as depicted in

the (figure 2.5). The light reflected from skin measured by the camera sensor depends on the distance from the light source to the skin tissue and the camera sensor. The variation of color measured by the camera sensor is mainly due to the motion-induced intensity/specular variation and pulse-induced subtle color changes.

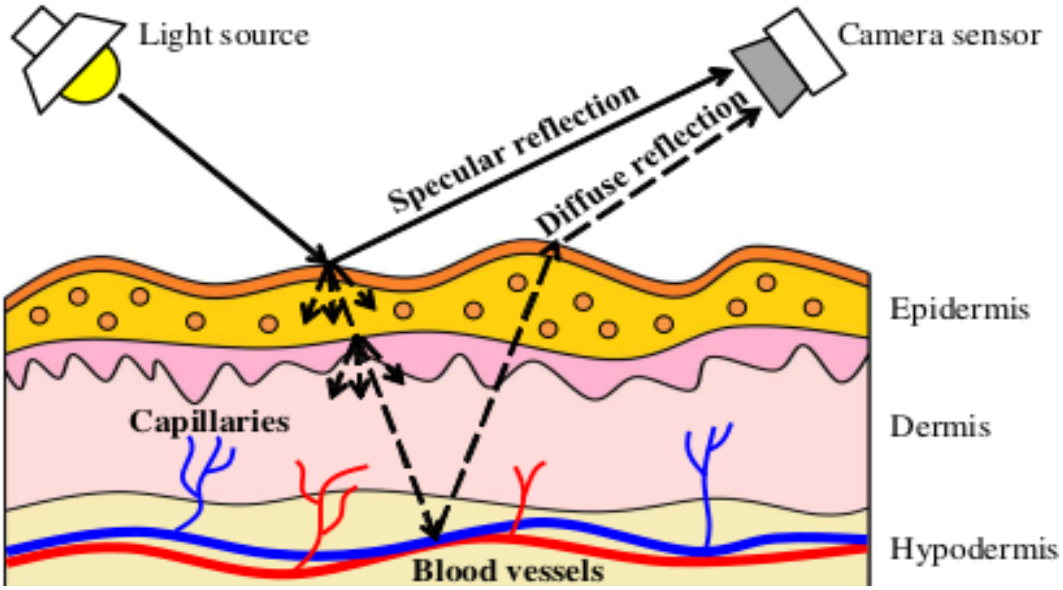

Figure 2.5: Skin Reflection Model Pictorial Presentation

According To the Skin Reflection Model, [10] the reflection of each pixel in a recorded image sequence can be defined as a time-varying function in RGB channels as

$$Ck(t) = I(t).(Vs(t) + Vd(t)) + Vn(t) \tag{2.1}$$

Here Ck(t) is a column vector that denotes RGB channels of the kth pixel, and I(t) denotes the luminance intensity level. It has a component of intensity variation because of distance changes between three components.
Vs(t) denotes the specular reflection, which is a mirror-like light reflection from the skin surface, and it does not contain any pulse information. We can write it as

$$Vs(t) = Us.(So + S(t)) \tag{2.2}$$

Here, Us denotes the unit color vector of the light spectrum while S0 and S(t) denote the stationary and varying parts simultaneously.

Vd(t) in (Eq 2.1) denotes the diffuse reflection, which means the emission of absorbed light

again to the camera sensor, and this component contains pulsatile information. We can write it as

$$Vd(t) = Ud.do + Up.p(t) \tag{2.3}$$

Here Ud is a unit color vector, up denotes the relative pulsatile strength and p(t) denotes the pulse signal, and Vn(t) in (Eq 2.1) denotes the quantization noise of the camera sensor.

Now by combining (Eq 2.1) ,(Eq 2.2) and (Eq 2.3) , we get

$$Ck(t) = I(t).(Us.(So + S(t)) + Ud.do + Up.p(t)) + Vn(t) \tag{2.4}$$

As Us, So, Ud, and do all are constant, so for the sake of simplification, it can be written as follows.

$$Uc.Co = Us.So + Ud.do \tag{2.5}$$

So our compressed Equation of Skin Reflection Model will look like

$$Ck(t) = Io.(1 + i(t)).(Uc.Co + Us.S(t) + Up.p(t)) + Vn(t) \tag{2.6}$$

Now the main task is to extract p(t) (knows as pulsating/pulse signal and contains information about HR) from Ck(t) in (Eq 2.6).

To make the problem simple and reduce the quantization error of the camera sensor, we will do spatial averaging over each frame, assuming sufficient amounts of skin pixels are in the frame. After applying spatial averaging and expanding the (Eq 2.6) we would get as follows

$$C(t) = Uc.Io.Co + Us.Io.S(t) + Up.Io.p(t) + Uc.Io.Co.i(t) + Us.Io.S(t).i(t) + Up.Io.p(t) \tag{2.7}$$

and now the approximating all AC modulation terms as zero because they will be minimal in magnitude than the DC terms thus, product terms can be neglected.

$$C(t) = Uc.Io.Co + Us.Io.S(t) + Up.Io.p(t) + Uc.Io.Co.i(t) \tag{2.8}$$

The (Eq 2.8),shows that the observation C(t) is a linear mixture of three source-signals i(t), s(t) and p(t). This implies that by using the linear projection, we are able to separate these source signals. Thus the task of extracting the pulse signal from the observed RGB- signals can be translated into defining a projection system to decompose C(t).

There are broadly two ways in which we can solve this signal demixing problem.

1. **Blind source separation based method - ICA(Independent component analysis)**
   The general procedure of the BSS-based PPG extracting method can be expressed as:

$$Y(t) = WC(t) \tag{2.9}$$

   where Y(t) denotes the factorized source signals consisting of the pulse and noise. W denotes the de-mixing matrix that can either be estimated by ICA.
   The BSS operation is followed by selecting the most periodic signal from Y(t) as the pulse. ICA have different limitations when estimating W:

   (a) PCA uses the covariance of RGB signals to estimate W (i.e., eigenvectors), which requires the variation in the amplitude of pulse and noise to be sufficiently different to determine the eigenvector directions.

   (b) ICA assumes that the components in Y(t) are statistically independent and non-Gaussian for deriving W, which requires C(t) to be a long signal to enable a statistical measurement. Furthermore, the procedure of BSS in estimating an exact W is completely blind (i.e., a black box), which is not tractable for algorithm development.

   Most importantly, BSS techniques are statistical and computational solutions for general signal-processing problems, which do not exploit the unique and characteristic skin reflection properties that can be used to solve the rPPG-specific problem. Especially illustrative in this respect is the ICA- based approach which normalizes the standard deviation of RGB signals upfront, thus ignoring the fact that the PPG signal induces different yet known relative amplitudes in the individual RGB channels [10].

2. **Model based method - POS(plane orthogonal to the skin):** In contrast to the BSS-based methods that impose no assumption on the colors associated with the source signals, the model-based methods use knowledge of the color vectors of the different components to control the de-mixing. We have used the model-based method for implementation and testing the system.
   To demix the RGB signal, we have followed the following process, which is known as POS Algorithm.

   (a) **Eliminate the dependence of C(t) on the average skin reflection color by temporal normalization** It can be done by dividing the RGB signal by their temporal mean, which will not affect the AC components, as i(t),S(t),p(t) are zero mean signals so mean of (Eq 2.8) is given by

$$u(C(t)) = Uc.Io.co \tag{2.10}$$

So, from this, we define a diagonal normalization matrix N

$$N.u(C(t)) = 1 \tag{2.11}$$

where 1 is an identity column vector, The Normalisation Matrix N will normalize the C(t) and give rise to Cn(t)

$$Cn(t) = 1.(1 + i(t)) + N.Us.Io.S(t) + N.Up.Io.p(t) \tag{2.12}$$

Now, as Cn(t) contains a constant signal having to mean 1, which is not useful in our work (remember we have to extract p(t)).

$$\tilde{C}n(t) = Cn(t) - 1 = 1.i(t) + N.Us.Io.s(t) + N.Up.Io.p(t) \tag{2.13}$$

The $\tilde{C}$n(t) is a zero-mean signal; now, we proceed further processing with it.

(b) Now, we will simply project $\tilde{C}$n(t) onto a plane orthogonal to 1, which is expressed as

$$S(t) = Pp.\tilde{C}n(t) \tag{2.14}$$

Here, Pp denotes a 2*3 projection matrix, and both the rows of this matrix are assumed to be orthogonal to each other.

Now, we need to define Pp. we will exploit the physiological properties of PPG absorption to define it. By using these properties, paper [10] found Pp to be a (2*3) matrix and is following.

$$Pp = [(01 - 1); (-21 - 1)], \tag{2.15}$$

Now we get S(t) rows as follows,

$$S1(t) = Gn(t) - Bn(t), S2(t) = Gn(t) + Bn(t) - 2Rn(t), \tag{2.16}$$

Now , we will extract final PPG signal h(t) from S(t) using alpha tuning[4], which can be expressed as,

$$h(t) = S1(t) + alpha.S2(t) \tag{2.17}$$

where

$$alpha = std(S1)/std(S2) \tag{2.18}$$

Signal h(t) is an rPPG signal in which we are interested, for our further processing.

Chapter 3

Methodology

3.1 System Design

In this section, we will present our workflow design of both of our final goals (Figure 3.1) and the design of the system which we have implemented till now (Figure 3.3).

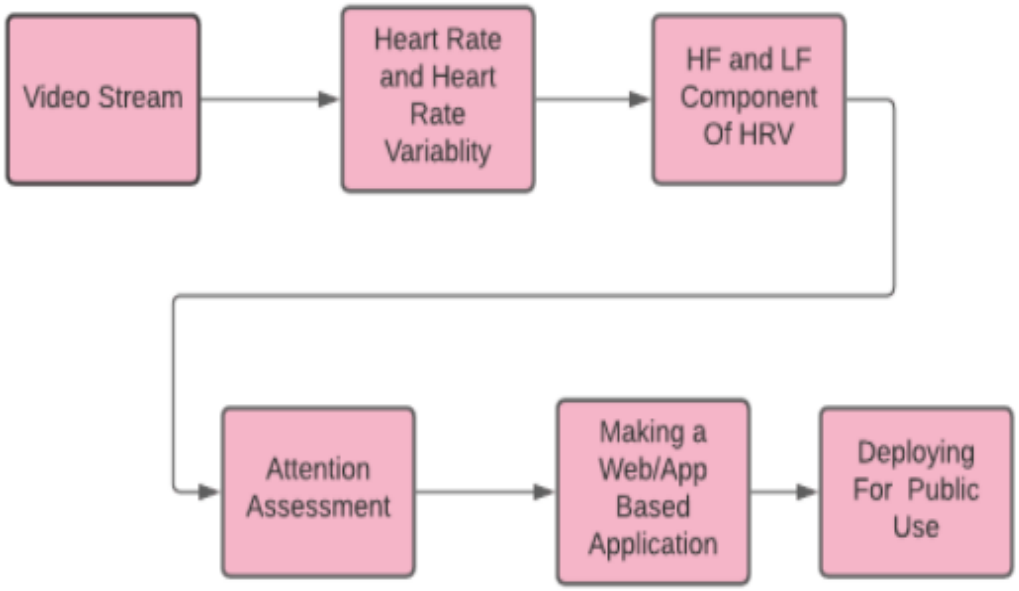

Figure 3.1: Flow Diagram Of Major Goals

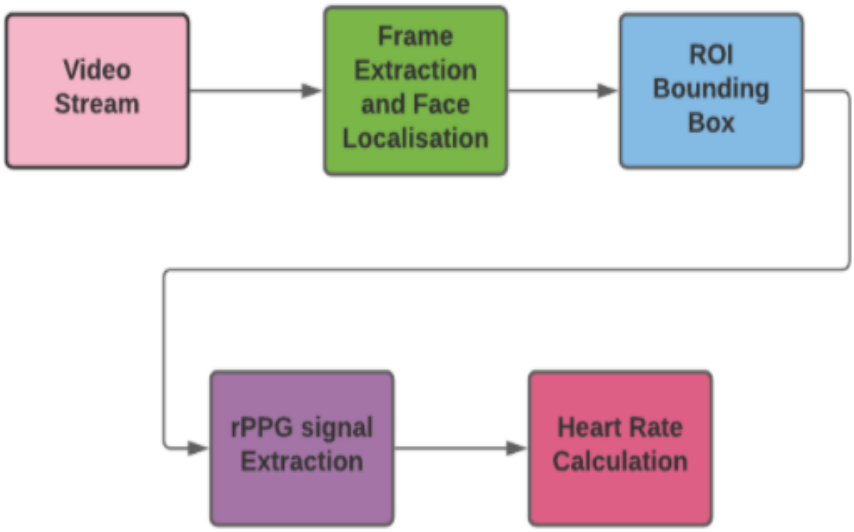

Figure 3.2: Flow Diagram of System Implemented

Flow Diagram 1 describing our final goals and objectives and Flow Diagram 2 describes the step we have taken till now for the measurement of Heart Rate. In the next section, we will discuss all the steps involved in a detailed manner.

3.2 Implementation

In this section, we will look into the details of the steps involved in the prediction of heart rate.

3.2.1 Video Stream

We have used the webcam of our laptop for this purpose and it streams the video at approx 30fps.

3.2.2 Frame Extraction

We have taken a sampling rate according to the FPS specification of the camera we have used so we don't miss any frame of video, because if we lose any frame then it will affect the measurement of heart rate and further calculation. Then we have simply sampled the video according to that fps to extract frames.

3.2.3 Face Localisation in the Frames

Face Localisation in the frames is one of the most important steps in the process of HR measurement. We have taken two paths for localizing the face as shown in the flow diagram below.

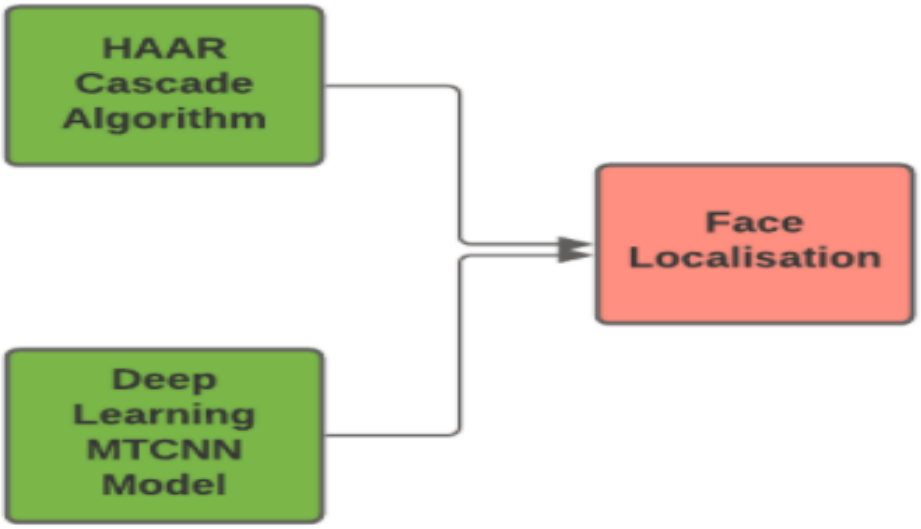

Figure 3.3: Path Taken For Face Localisation

In the first trial, we have used the OpenCV Cascade Classifier for this purpose. The algorithm of this library of OpenCV is based on the basic concept of the HAAR Cascade Algorithm. In the second trial, we have used the pre-trained MTCNN library which is based on the concept of Deep Learning.

1. **HAAR Cascade Algorithm(Viola-Jones algorithm):** The Viola-Jones algorithm [4](also known as Haar cascades) is the most common algorithm in the computer vision field used for face detection on the image.
   In this algorithm, the image to be used is divided into different kinds of sub-windows and multiple Haar-like features to compute it at different scales and positions for each sub-window is Used. The main features are selected using the Adaboost algorithm. Then each sub-window is checked for the presence or absence of face using a cascade of classifiers.
   The Harr features are of different types and is used for different purposes some of them are

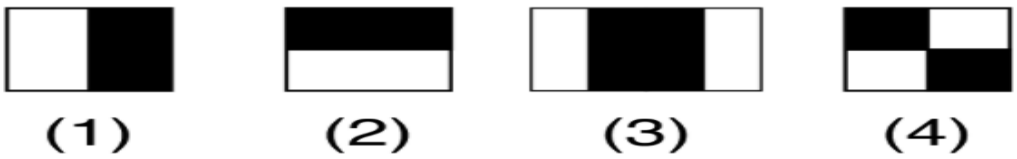

Figure 3.4: HAAR Features

Features = sum(pixels in the black area) - sum(pixels in the white area)

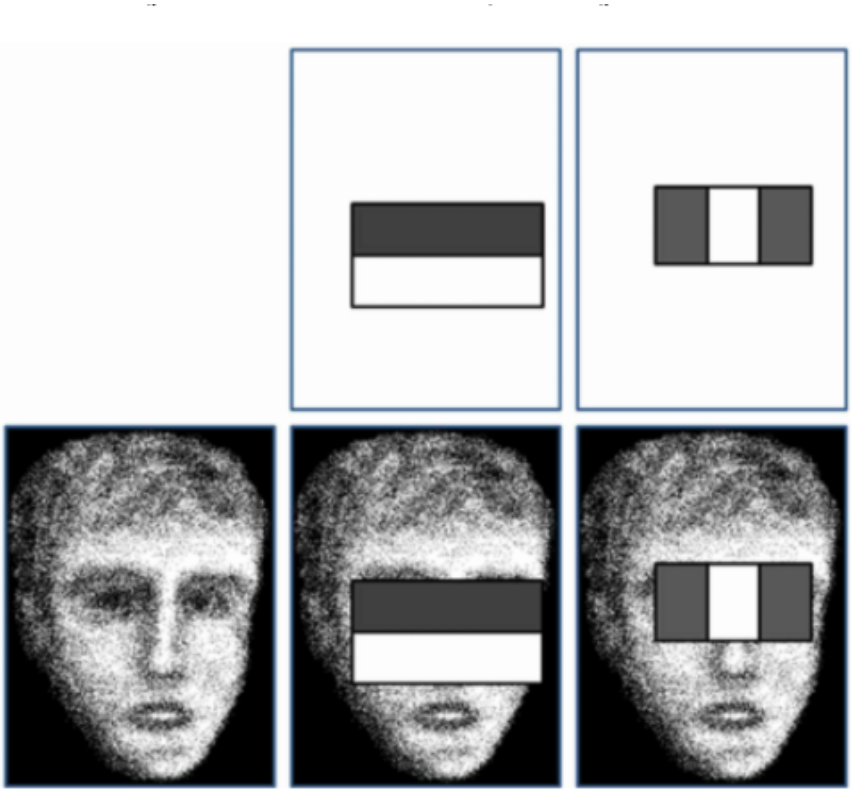

Figure 3.5: Face Localisation Using Haar Features

In the (Figure 3.5) above, when we calculate the second image features, it will give more features because the bridge is lighter than the nearby area. But if the same features, we keep on the head of the face, we will get fewer features. In the third image, we also get more features as it can detect the eye region since the eye region is darker as compared to the region below it.

It should be noted that only a single feature is not capable of detecting faces with high accuracy. But, when many such features vote for the presence or absence of a face, the detection becomes very accurate and robust.
These features have actual real importance in the context of face detection:

(a) Eye regions tend to be darker than cheek regions.

(b) The nose region has more bright pixels than the eye region.

Therefore from the above given four rectangles along with the corresponding difference of sums, we are able to get the features that can classify the face. To detect which features belong to a face from the available number of features we use the Ada-Boost algorithm to select which ones correspond to facial regions of an image.

As we can imagine, using the above rectangle fixed sliding window on the image across every (x, y) coordinate of an image, followed by computing all those features using Haar to classify the face features. This whole process can be computationally expensive.

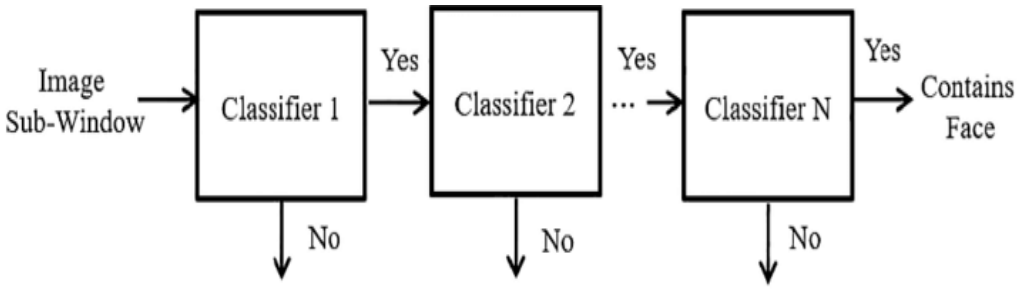

Figure 3.6: Work Flow Of Viola Jones Algorithm

To overcome this, Viola and Jones introduced the cascade concept. At each stop along the sliding window path, the window must pass a series of tests where each subsequent test is more computationally expensive than the previous one. If any of the tests fail, the window is automatically discarded and this saves lots of time and finally, the Viola-Jones algorithm is able to detect objects in real-time.

2. **Deep Learning-Based Approach:** In this approach we have the MTCNN(Multi-task Cascaded Neural Network) method of Deep Learning. MTCNN detects faces and facial landmarks on images/videos. This method was proposed by Kaipeng Zhang et al. in their paper[11].

   The whole concept of MTCNN can be explained in three stages out of which, in the third stage, facial detection and facial landmarks are performed simultaneously. These stages consist of various CNN's with varying complexities. A pictorial View of Various steps involved in MTCNN is shown below in (Figure 3.7). A simpler explanation of the three stages of MTCNN can be as follows :

   (a) In the first stage the MTCNN creates multiple frames which scan through the entire image starting from the top left corner and eventually progressing towards the bottom right corner. The information retrieval process is called P-Net(Proposal Net) which is a shallow, fully connected CNN.

   (b) In the second stage all the information from P-Net is used as an input for the next layer of CNN called as R-Net(Refinement Network), a fully connected, complex CNN which rejects a majority of the frames that do not contain faces.

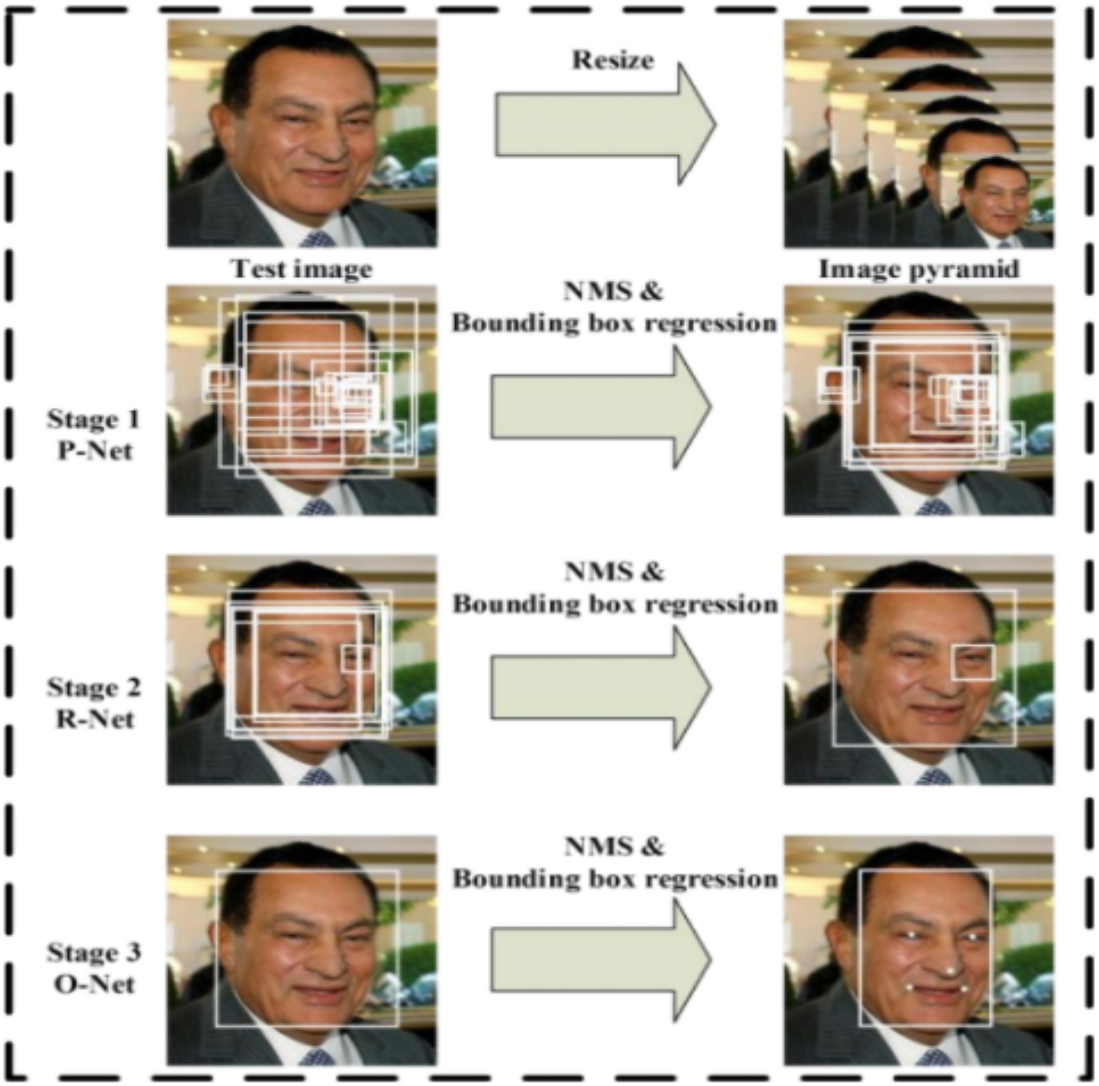

Figure 3.7: Pictorial Presentation Of Steps Involved in MTCNN

(c) In the third and final stage, a more powerful and complex CNN, known as O-Net(Output Network), which as the name suggests, outputs the facial landmark position detecting a face from the given image/video

Both of these algorithms are very fast and accurate in finding faces in the frames. Still, We further proceed with the Deep Learning MTCNN approach because in the Haar Cascade algorithm approach sometimes algorithms recognize the faces of statue and Images in the background which affect the further calculation but there is no such issue in MTCNN.

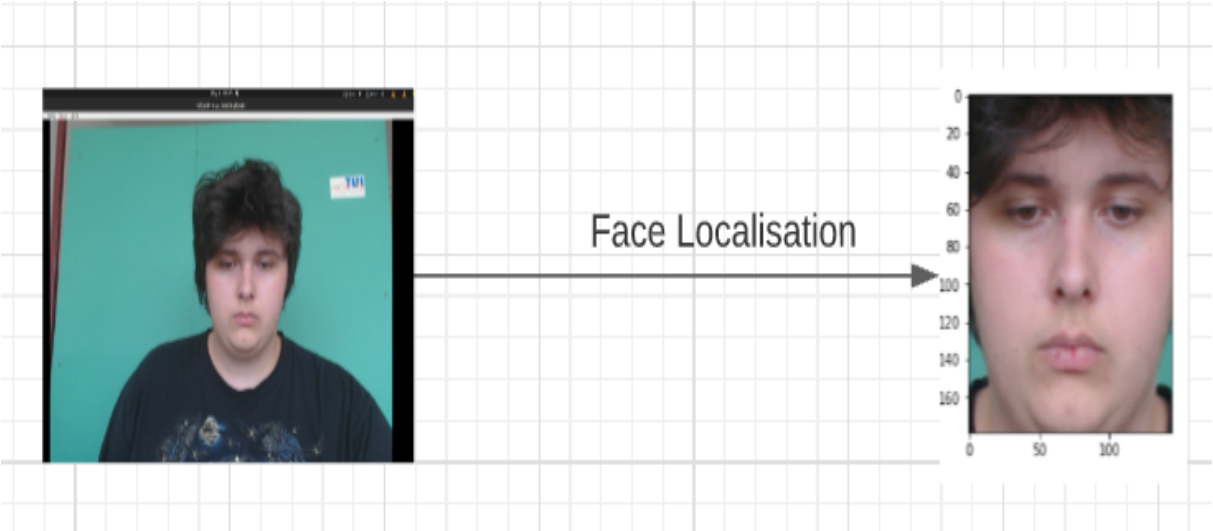

Figure 3.8: Localised Face in Frame using MTCNN

3.2.4 ROI Selection

To select the ROI on which we have to work further we have explored many paths but the most prominent are shown in the figure below.

1. **Using Full Face :**In this we have used the full face of the subject for further processing, we haven't removed any part of the face like eyes or hair.
2. **Using Forehead Part:** In this we have empirically founded a bounding box which gives us the pixel of forehead region in the frames of videos as in [7].
3. **Using Full Face after Removal of Eyes, Hair and Low illumination Region:**In this we have used the OTSU segmentation algorithm which removes all low illumination regions, eyes, and hairs.

   The main steps involved in the OTSU algorithm according to [6] are following

   (a) Divide the image into two parts using every grayscale value (0-255) as a potential threshold.

   (b) Computing the sum of the weighted variances of the two parts formed is minimum or not. That is Wo*Vo+W1*V1.
   where Wo, W1=No. of pixels in the first and second part Vo, V1 =Variance of the first and second part

   (c) The threshold at which the sum is minimum is called the Otsu's threshold.

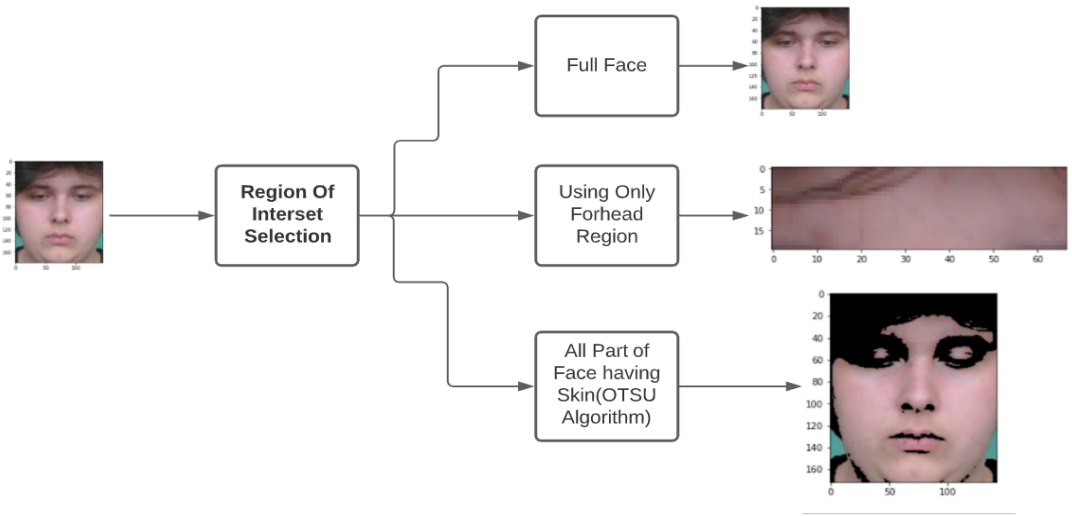

Figure 3.9: Region Of Interest Selection

When we have used full-face as ROI then the signal received from the face is very noisy, and we are not able to retrieve the actual signal, and so we may get confused while finding the peak of signal for heart rate calculation as clear from the figure shown below.

Similarly, when we use the forehead region of subjects as ROI, then results aren't looking good because sometimes it's tough to keep hair and eyebrows of subjects out of ROI, which leads to extracted RGB and rPPG signal very noisy and hence gives high error rates so this method we don't find to be promising.

Now we have explored a new ROI that subjects full face but after removing hair, eyebrows, and low illumination parts of the Face. The result obtained from using full face with eye, hair, and low illumination region removed is giving good results and promising, so we have decided to move further with this ROI selection method as the main method and used the other two for comparison.

3.2.5 rPPG Signal Extraction and Heart Rate Calculation:

The steps involved in the extraction of rPPG signal and so HR is described in the (figure 3.10). From the discussion in the last subsection, we have Frames of the video with ROI parts only now we extracted the pixels of that frames in terms of RGB channels.

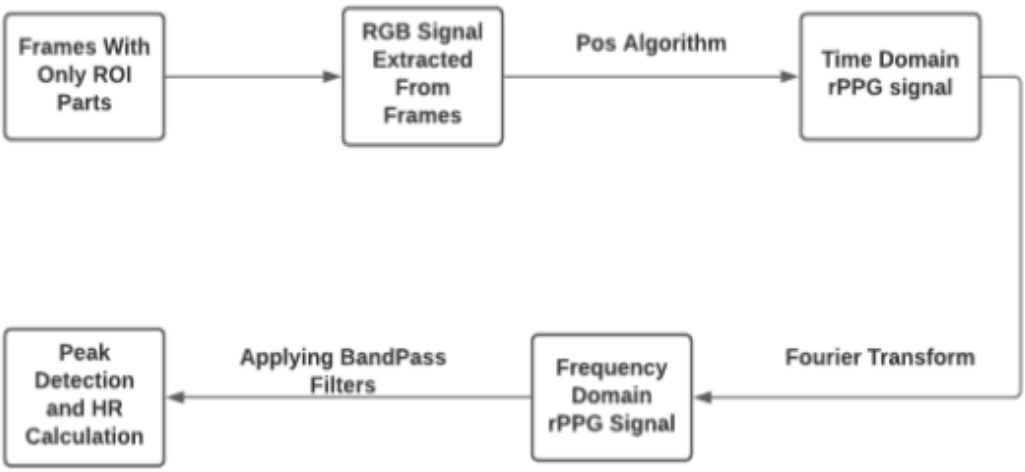

Figure 3.10: Flow Diagram of rPPG Extraction

From the skin reflection model (Discussed in section 2.2.3), we know that the raw RGB channels contain so much noise and other unwanted constant signals which have no meaning for our work, and hence they must be filtered out to get the signal which has meaning in context.

Now we also know that reflection from skin can be modeled using equation 3.1

$$C(t) = i(t)(Uc * Co + Us * S(t)) + Up.p(t) \tag{3.1}$$

In the above, equation C(t) is raw RGB signal, and p(t) denotes the pulse signal; other variables have the same meaning as discussed in the Literature review under section 2.2.3.

Now we have extracted the pulse signal p(t) using the POS algorithm whose steps were discussed under section 2.2.3. The output of the POS algorithm is rPPG signal.

Signal filtering : The rPPG signal obtained from the POS method still contains some noise and unwanted signal, which must be filtered out to get Heart rate. The next step of the process is to convert the time domain rPPG signal into frequency domain signal so filtering can be carried out.

After carrying out the domain conversion(time to frequency) we will first filter out all the signals which have frequencies outside of the human heart range (0.7-4 Hz/42-240 bpm) from the spectra.

Now We need to find the frequency corresponding to the peak of the signal obtained after filtering. This frequency corresponds to the Heart Rate of the window we have selected for calculation.

How we made the Heart Rate calculation system Live: As we can see that all the process which we have discussed and implemented till now depends on its previous process completed and hence they all are sequential in nature at the same time the Library used to capture the live stream using the webcam of the laptop is blocking in nature which means it will not allow other processes to go hand in hand it must so it is not allowing us to make our system live.

To solve this issue we have used the concept of Multi-Threading which gives us the flexibility of parallel computing. Now we are capturing and storing frames using one thread and doing all the processing in another frame. This makes our system very fast and robust, so we have achieved our goal of making the system live.

Chapter 4

Evaluation and Result Analysis

In this chapter, we will discuss the dataset used for validating our model, the evaluation strategy, and finally the results we got in testing.

We have used a public dataset(can be accessed only after permission from BioBank) available here, We have given it name Dataset1. This dataset contains videos of students doing some task like solving a maths problem and the two signals for each subject (recorded while performing a task), one is an ECG signal and the other is the Heart Rate signal.

Details About Dataset

1. Frame rate of the video is 30 frames per second(fps).
2. Both the signals(PPG and Heart rate) also have a sampling rate of 30.

Assumption During Testing

1. Subjects should be a good lighting conditions like natural light.
2. Subjects should not be moving suddenly during the whole testing.

Apart from this dataset, We have tested our model on a Dataset created by us, we have given it the name Dataset2. To make this dataset we have recorded video using the WebCam of the laptop (Lenovo Ideapad 330 having WebCam which can record/stream video at 30 FPS) and at the same time, we have used a Pulse Oximeter (ISO Certified) to record the Heart Rate. But this dataset contains videos of only single participants because the current situation is not permitting us to do a large extent of testing with a large number of participants.

The figures 4.1, 4.2 and 4.3 show the rPPG signal in the frequency domain obtained from applying different ROIs selection methods on the subject video of Dataset1.

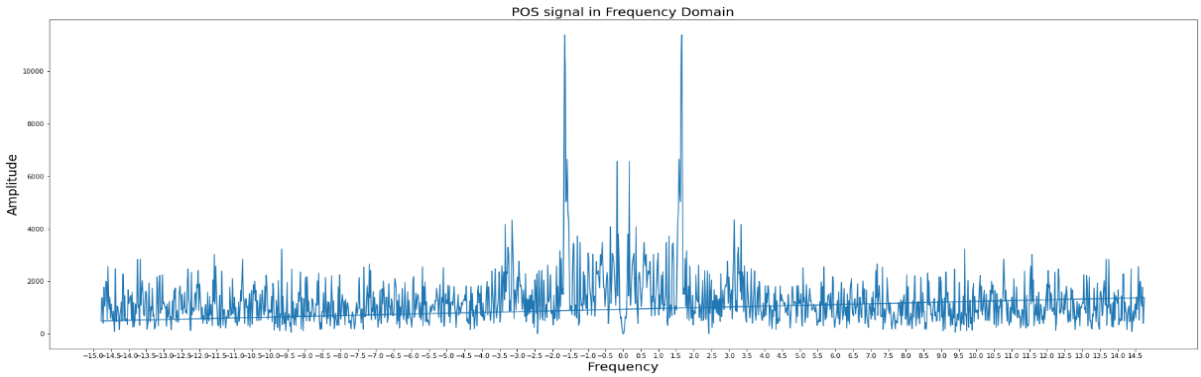

Figure 4.1: Frequency Domain rPPG Signal When ROI is Forehead Only (Dataset1)

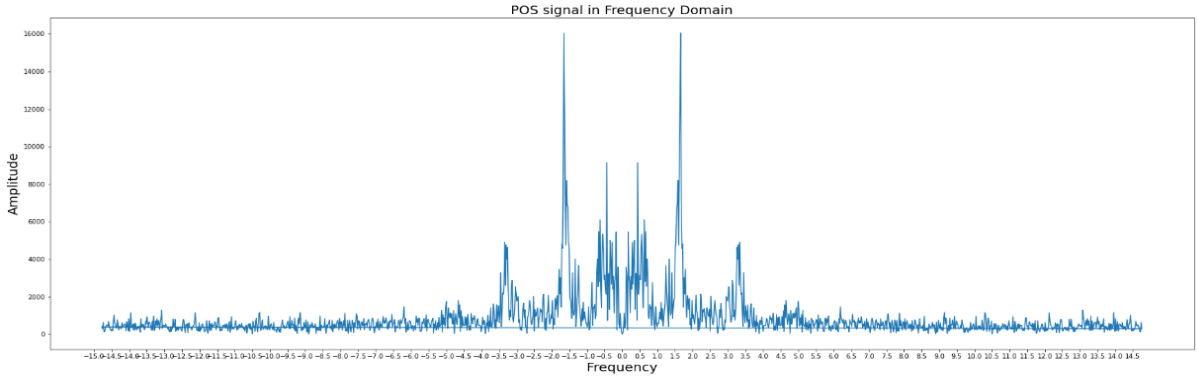

Figure 4.2: Frequency Domain rPPG Signal When ROI is Whole Face (Dataset1)

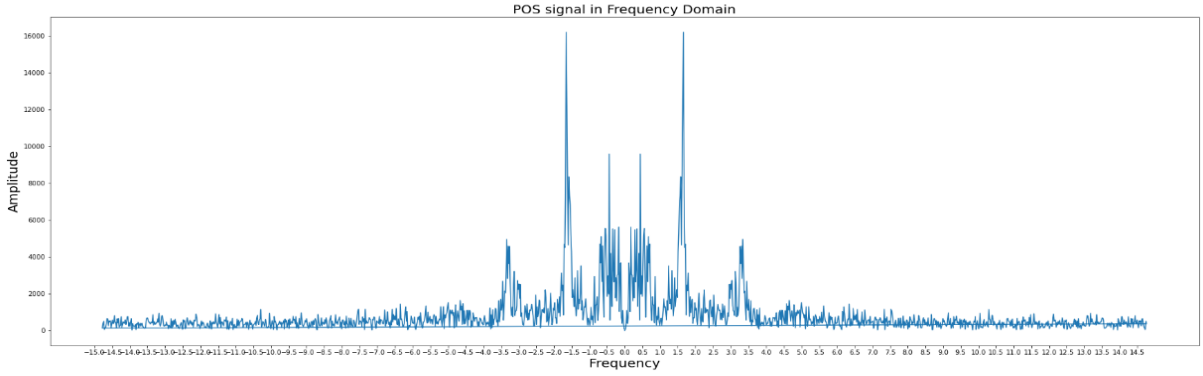

Figure 4.3: Frequency Domain rPPG Signal When ROI is Whole Face (With OTSU filter Applied) (Dataset1)

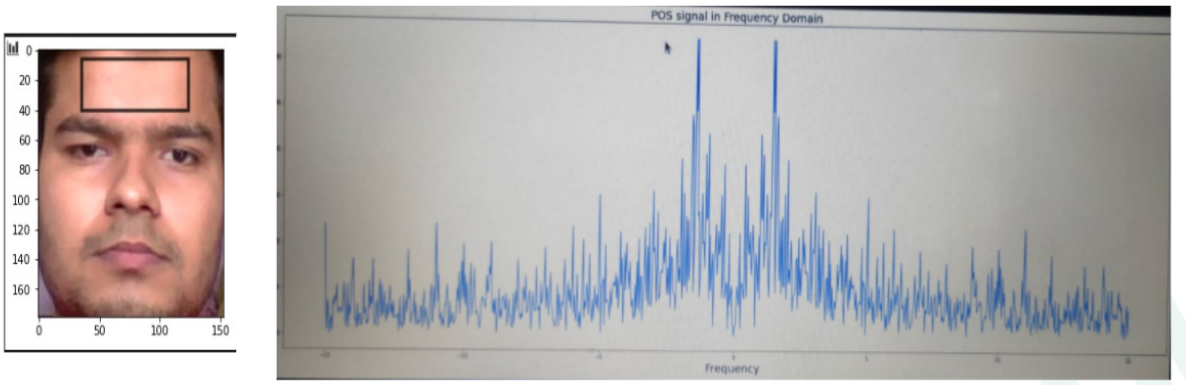

Figure 4.4: Frequency Domain rPPG Signal When ROI is Forehead Only (Dataset2)

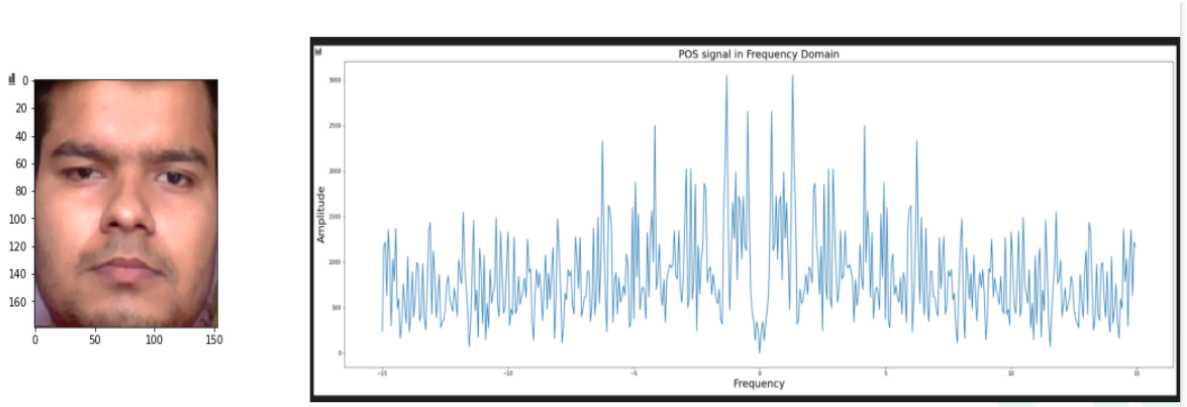

Figure 4.5: Frequency Domain rPPG Signal When ROI is Whole Face (Dataset2)

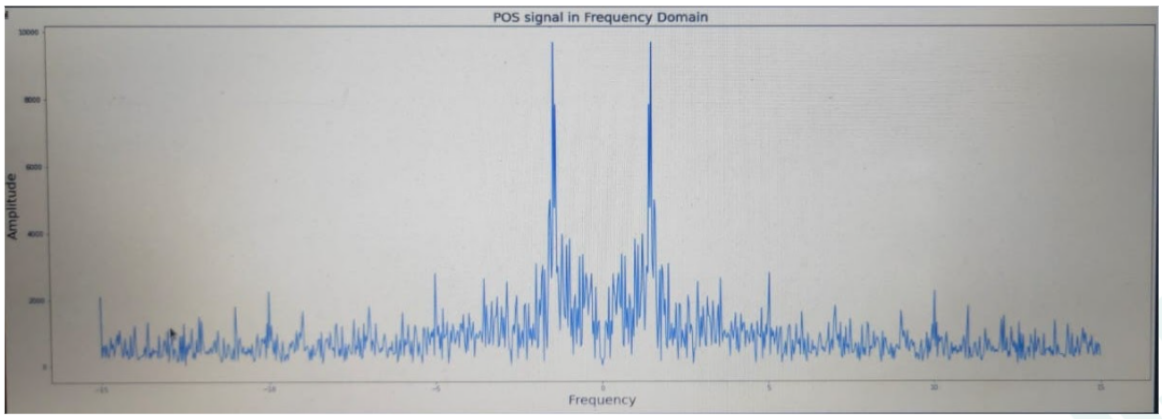

Figure 4.6: Frequency Domain rPPG Signal When ROI is Whole Face (With OTSU filter Applied) (Dataset2)

The figures 4.4, 4.5 and 4.6 show the rPPG signal in the frequency domain obtained from applying different ROIs selection methods on the subject video of Dataset2.

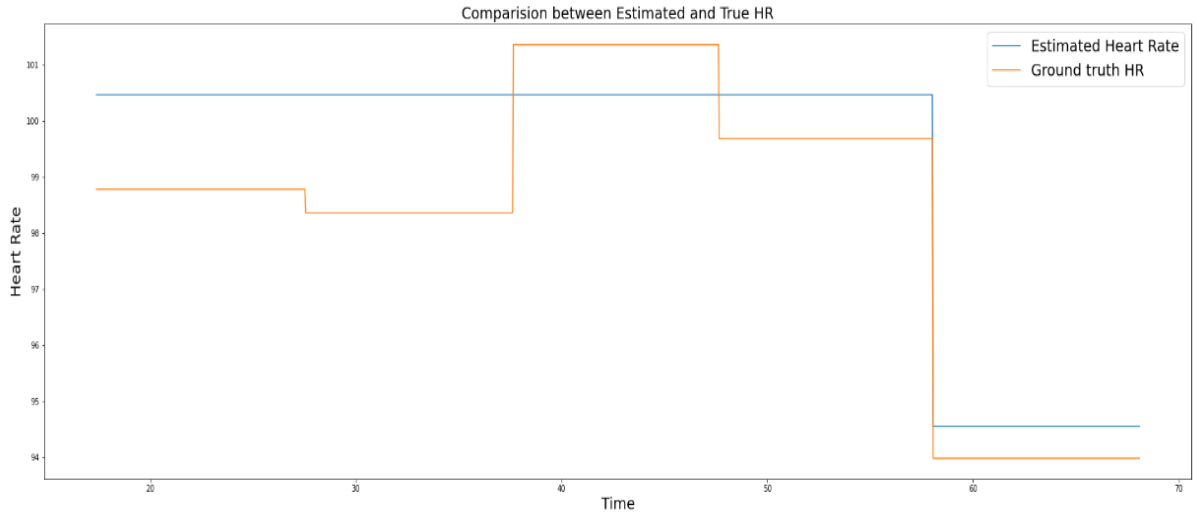

Figure 4.7: Comparision of Heart Rates of Ground Truth and Estimated Signals

We can see that the only ROI selection via OTSU algorithm is working well on Dataset2 (figure 4.6) other methods gives noisy signal in which finding a peak related to heart rate is a challenging task, which we have discussed in the implementation part also and these results verify that.

Figure 4.7 shows the Heart Rate Plot of the Ground Truth Signal and measured Signal on one subject.

We have calculated the mean absolute error between the True Heart Rate and Estimated Heart Rate and Table 4.1 below shows that.

Subjects	For Whole Face	Forehead Only	Using OTSU
Sub 1	1.20	1.54	1.20
Sub 2	5.81	4.12	4.12
Sub 3	2.72	2.72	2.72
Sub 4	4.83	3.01	3.01
Sub 5	3.07	3.07	2.89
Average Error	3.526	2.892	2.78

Table 4.1: Comparision of mean absolute error between the True Heart Rate andEstimated Heart Rate for different ROIs

The above-average error values for different ROIs show that the OTSU ROI selection method is giving the best results for Heart Rate (-/+ 3) to the value obtained from ground truth data. The reason for OTSU performing better than others is that it does not consider the pixels which are not well illuminated and it also removes pixels containing the facial hair from the face.

Chapter 5

Plan For Future Work

The future plan of our work can be summarised in the following points.

5.1 Testing Heart Rate Prediction System

Testing our Heart Rate Prediction System on a large set of subjects having different skin tones so we can validate our system well.

5.2 Calculating HRV and Predicting Human Attention

As we have implemented the heart rate measurement system successfully, our next goal is to find the HRV (which is not a major task now) and then use the features of HRV(HF and LF components As discussed in the literature review section) to predict the Attention Level.

5.3 Testing Our Attention Prediction system

Once we are ready with the Attention Assessment System, then our next goal will be to do a rigorous test of our attention prediction system using a sufficiently large pool of participants so we can validate our system for real-life use.

5.4 Making Web-Based Application

Once the testing of the system is finished, then our next goal will be to make some web-based extension that can integrate with online meeting applications like google meet to predict the attention level of the participants. We want this to integrate with google meet because we want to develop a system that can provide a tool to teachers/mentors by using which they can see the attention level of students, which is currently not possible in an online mode, unlike physical classes where they use to see faces of every student and observe whether they are grasping things

or not. This will facilitate online learning in a smooth way and increase the overall efficiency of online learning.

5.5 Other Scopes

1. Another area where we can find opportunity is in telemedicine as we can see as the covid cases are rising up the price of a pulse oximeter is shooting up, so if we are able to predict the saturation level of oxygen with good accuracy as we have achieved in predicting heart rate, then we can make a mobile app-based solution to replace the pulse oximeter which will be cost-effective and easily available to use.

2. Once we will be ready with the Attention Assessment System, then We can also develop a web/App-based personal tool that a person can use while working and then analyze his/her attention level while working. This will give him an idea about his working ability so he can make the best out of himself.

Bibliography

[1] A Artifice, J Sarraipa, and R Jardim-Goncalves. "Methodology for Attention Detection based on Heart Rate Variability". In: *2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES)*. IEEE. 2018, pp. 000395–000400.

[2] Chih-Ming Chen, Jung-Ying Wang, and Chih-Ming Yu. "Assessing the attention levels of students by using a novel attention aware system based on brainwave signals". In: *British Journal of Educational Technology* 48.2 (2017), pp. 348–369.

[3] Diana K Darnell and Paul A Krieg. "Student engagement, assessed using heart rate, shows no reset following active learning sessions in lectures". In: *PloS one* 14.12 (2019), e0225709.

[4] Theo Ephraim, Tristan Himmelman, and Kaleem Siddiqi. "Real-time viola-jones face detection in a web browser". In: *2009 Canadian Conference on Computer and Robot Vision*. IEEE. 2009, pp. 321–328.

[5] Wan-Hua Lin et al. "Comparison of heart rate variability from PPG with that from ECG". In: *The international conference on health informatics*. Springer. 2014, pp. 213–215.

[6] Dongju Liu and Jian Yu. "Otsu method and K-means". In: *2009 Ninth International Conference on Hybrid Intelligent Systems*. Vol. 1. IEEE. 2009, pp. 344–349.

[7] Carmen Nadrag, Vlad Poenaru, and George Suciu. "Heart Rate Measurement Using Face Detection in Video". In: *2018 international conference on communications (COMM)*. IEEE. 2018, pp. 131–134.

[8] Taishi Nagasawa and Hiroshi Hagiwara. "Workload induces changes in hemodynamics, respiratory rate and heart rate variability". In: *2016 IEEE 16th international conference on bioinformatics and bioengineering (BIBE)*. IEEE. 2016, pp. 176–181.

[9] Maria Dolores Coca Peláez et al. "Photoplethysmographic Waveform Versus Heart Rate Variability to Identify Low-Stress States: Attention Test". In: *IEEE journal of biomedical and health informatics* 23.5 (2018), pp. 1940–1951.

[10] Wenjin Wang et al. "Algorithmic principles of remote PPG". In: *IEEE Transactions on Biomedical Engineering* 64.7 (2016), pp. 1479–1491.

[11] Kaipeng Zhang et al. "Joint face detection and alignment using multitask cascaded convolutional networks". In: *IEEE Signal Processing Letters* 23.10 (2016), pp. 1499–1503.