Emotion Recognition Using Wireless Signals : Remote Photoplethysmography and KNN Classifier

Jui Mhatre

Abstract—Emotion Recognition plays an important role in understanding human behavior. It finds its utility in various domains such as healthcare, automobile industries, understanding social interactions, fraud detection, and many more. Analyzing a person's emotions in a controlled environment with various devices has been challenging since it adds to human anxiety, which manipulates the readings. This presents a need to devise ways to recognize and study emotions in a wireless manner. There have been previous works done to predict emotions by obtaining photos of subjects whose facial features were analyzed to identify emotions. Though this is an easier way of obtaining results, the data obtained could be wrong since such features could be faked by people. Physiological signals play an important role here. But obtaining physiological signals like Electrocardiography (ECG), Electroencephalography (EEG), Heart Rate Variability (HRV), Blood Volume Pulse (BVP), Galvanic Skin Response (GSR), Electrodermal analysis (EDA), etc. require dedicated equipment, handling skills, and its knowledge. Moreover, human emotions are contaminated when devices are planted on the body since it produces discomfort for them and thus produces incorrect signals. Both physical and physiological methodologies have their advantages and disadvantages. Taking best from both worlds, this paper proposes a technique for emotion recognition that combines both approaches where both physical and physiological signals are used. We use Remote Photoplethysmography (rPPG) to identify the HRV and HR signals which are analysed and classified for emotion recognition. RAVDEES [1] dataset is used for obtaining HRV and HR signals using rPPG. We further trained the KNN classifier on obtained HRV signals to identify emotions and obtained 40.45% accuracy for diverse set of 8 emotions compared to state-of-art model which gives 40.49% accuracy for 4 emotions using pulse rate variability (PRV). We also verified that HRV signals are enough to recognize emotions using SWELL and WESAD dataset unlike previous work which used other signals in conjunction with HRV to recognize emotions. We use KNN-classifier for training emotion recognition on combined SWELL and WESAD [2] [3] datasets. The proposed classifier with SMOTETomek sampling gives improved results for imbalanced and overlapping dataset. We have done comparative study for Logistic regression (accuracy = 50.34%), Support Vector Machine (SVM) classifier with OVR strategy (accuracy = 85.36%), K-nearest-neighbour (KNN) with k =500 (accuracy = 93.27%) and Naive Bayesian classifier (accuracy = 35.46%).

Index Terms—Remote Photoplethysmography, Independent Component Analysis, Wireless, Classification, Undersampling, Oversampling, Imbalance, Overlap, SMOTETomek, KNN-Classifier, Support Vector Machine, Naive Bayesian classifier.

1 Introduction

It has been established that emotions have a huge impact on physiological signals directly [4] [5]. Any change in emotions shows a change in physiological signals. Physiological signals are very useful because, as compared to other signals, they cannot be controlled by people themselves. It is highly impossible to suppress or mask these biosignals which represent emotions [6]. These signals have become a strong alternative to facial expressions, gestures, and vocal traits. The multimodal approach of combining different physiological signals from biosensors like ECG, an electroencephalogram (EEG), an electromyogram (EMG), electrodermal activity (EDA) or galvanic skin response (GSR), a photoplethysmogram (PPG), or blood volume pressure (BVP), or a respiratory inductive plethysmograph (RIP) is used.

Emotions directly impact our heart rate. It is known that the autonomic nervous system (ANS) influences our heart rate. It consists of the sympathetic and parasympathetic nervous systems. The sympathetic nervous system is activated in case of exciting emotions like fear, anxiety, stress, frustration and increases the heart rate. The parasympathetic nervous system is activated in relaxed situations like being happy, compassionate, calm, and decreases the heart rate. The heart rate is the number of heartbeats per minute. Heart rate variability (HRV) is the fluctuation in the time intervals between adjacent heartbeats. HRV analysis shows its neuro-cardiac function which is generated by heart and brain interaction and non-linear relation with ANS [7]. Studies are done in [8] to verify whether Heart Rate Variability is a very good and adequate measure for emotion recognition. It concludes that although HRV is currently used sporadically in experiments that are designed to assess human emotional states, no research has yet verified its validity as a tool for assessing human emotion. Heart rate variability (HRV) analysis is considered a noninvasive technique for the assessment of the balance between sympathetic and parasympathetic branches of the ANS by its spectral analysis and it has been proposed for human emotion recognition [9], [10]. HRV has time domain and frequency domain attributes. Time-domain attributes pertain to inter-beat interval measurements of the heart. It is the measurement of time intervals between two consecutive heartbeats. Some of them include SDRR (standard deviation of RR intervals), RMSSD (root mean square of successive RR interval differences), pNNx (Percentage of successive RR intervals that differ by more than x ms).

The extraction of heart rate is done through various mediums, from high-end Heart Rate Variability monitors which are used in specialized environments to easily available wearable devices. It is usually obtained by ECG signal preprocessing. PPG, Photoplethysmography is another widely used technique for measuring HRV where wearable devices emit light to the skin which is reflected back. The amount of reflection changes due to changes in blood volume caused by constriction and dilation of capillary tubes carrying blood. rPPG is an emerging technique to measure HRV remotely. It works on the same principle as PPG and captures Red, Blue, and Green light reflections from the skin using cameras and webcams to obtain pictures or videos of subjects. This paper focuses on wireless emotion recognition and is done in 2 major phases. Phase I includes using the rPPG technique to find HRV where spectral analysis is done to obtain BVP signals which are processed to obtain HRV. In Phase II, obtained HRV signals are analyzed to recognize emotions using the pre-trained model. In order to train the recognizer model, we use various classification techniques.

The paper is organized as follows: Section 2 presents materials and methods used in the study, Section 3 gives details of the proposed method. Section 4 gives details of how the proposal was tested and experiments were carried out. Section 5 and 6 discuss the results and provides conclusive remarks.

2 Related work

Current research of emotion recognition has a rich history traced back to the 1800s with the publishing of Charles Darwin's The Expression of the Emotions in Man and Animals (originally published in 1872) and G.G. Duchenne de Bologne's The Mechanism of Human Facial Expression (originally published in 1862) [11]. Since then emotion recognition has been a field of research. Earlier facial expression features, gestures were used to find emotions. With advancement of biomedical devices and establishment of relation of emotions and physiological signals, research is done to identify emotions by analysing physiological signals. Some recent work is done in this field,

- Reference [12] shows emotion recognition using Cohn–Kanade (CK +) and Japanese Female Facial Expression (JAFFE) datasets. It involves facial expression features extraction from photos in dataset. Deep Neural networks(DNN) is trained using these datasets which gives accuracy of 95.24%. This model gives improved results, but a major disadvantage using this model is it uses physical signals which are not reliable.
- Combination of physical and physiological signals is used in paper [13]. The multimodal approach is used to do classification where input signals are electroencephalogram and facial expression. The stimuli are based on a subset of movie clips that correspond to four specific areas of valance-arousal emotional space. Facial images are analyzed using neural network and SVM classifier classifies emotions in EEG signals. Using a SVM classifier, the results show that the accuracies of multimodal fusion detections as 82.75%, which were higher than that of individual classification using facial expression (74.38%) or EEG detection (66.88%).
- Just physiological signals are used in [14] for emotion recognition. This paper uses ECG -based emotion recognition algorithm for human emotion recognition. Music is used to induce emotions and ECG is measured which is passed to the LS-SVM classifier. The results show that the correct classification rates for positive/negative valence, high/low arousal, and four types of emotion classification tasks are 82.78%, 72.91%, and 61.52%, respectively.
- ECG-based classification using AMIGO dataset in [15] gives better results by using wavelet transform for signal analysis.In this paper, when using two dimensions simultaneously (four classes), the performance of the KNN classifier dropped, possibly due to the similarity or low distance between features of each class. On the other side, the Ensemble Tree classifier accuracy was similar to that achieved in one-dimension classification, possibly due to the classifier being deep enough to capture the differences between classes of the scattering features of each signal. PCA can be used to reduce the dimensionality of the input vector, but it adds an additional step in the algorithm and tends to degrade classifier performance. This paper uses ECG data and increased performance of 90.2% accuracy.
- Eye fixation data was used as the emotional-relevant feature in this investigation in [16]. The eye-tracking data was collected and recorded using an add-on eye-tracker in the VR headset. Three classifiers were used in the experiment, which are k-nearest neighbor (KNN), random forest (RF), and support vector machine (SVM). Comparative results showed that random forest gave the highest accuracy of 80.55%. It is a novel approach since there are no previous studies on emotion recognition based on eye fixation's position purely in VR stimuli. These classifications are based on data collected using devices.
- Photoplethysmography is used to collect physiological signals in [17]. PPG device can be used to get Heart Rate Variability signals. However, the quality of these signals (in terms of added disturbances) could be not always optimal, since they are susceptible to many factors, e.g. motion artifacts, ambient light, the pressure of contact, skin color, and conditions. Therefore, methods for artifacts correction play a pivotal role and consequently influence the results [17]. The studies show that HRV is a good measure for emotion recognition. They have used the SVM classifier. Further improvement is seen when EDA signals were used in addition to HRV. This model gives 66.67% accuracy.
- HRV data analysis is used for emotion recognition in [18] and [19]. [18] uses a hybrid signal optimization approach using radar and camera to remove the influence of body motion and light conditions on the physiological signal. The proposed system could achieve high classification accuracy of 89.6% for 10-fold cross-validation at sample level, and 71.0% for cross-validation at subject level. It uses 2 physiological signals, respiratory and heart rate.
- Paper [19] compares static and dynamic subjects.It uses XGBoost classifier for emotion recognition. Using the boosting methods for classification for the quasi-dynamic situation, we obtain about 80% accuracy in comparison to the 90% accuracy for the static situation. The result is significant and confirms the feasibility of the solution. Metrics such as recall (79 %), precision (80 %) and Macro-F1 (73.9 %), detect the imbalances in classes.
- Due to inability of all people to use medical appliances to measure physiological signals, there arose a need to read

these signals in wireless manner. Remote Photoplethysmography is the answer to it. [20] uses datasets which have signals read using rPPG technique. CAS(ME)^2 dataset is used in this work. In this work, the PRV is obtained using remote photoplethysmography. They prove that from a simple RGB camera, it is possible to assess the emotional state of a person by analysing their pulse rate variations. This optimistic finding is supported by surprising results and an accuracy rate of around 60the CAS(ME)2 dataset using SVM classifier.

Remote Photoplethysmography is a largely used technique of computer vision in various fields of medical sciences. Heart Rate [21] [22] , Pulse waveform generation are some of the applications of rPPG [23]. Earlier rPPG used hours of videos to extract HRV but in [] experiments focus on obtaining ultra-short HRV measure as a proof-of-concept/technology demonstrator for longer duration applications. This method extracts a clean BVP signal from the input via a two-step wide and narrowband frequency filter to accurately time heartbeats and estimates heart rate variability. Moreover, it does not require any rPPG specific training and it can perform its analysis with real-time speeds. It performs an in-depth HR and HRV evaluation on an exhaustive collection of 13 public and self-recorded datasets exploring a varied range of unique facets. Most recent works have applied deep learning to extract either heart rate or the blood volume pulse directly from camera images. They rely on the ability of deep networks to learn which areas in the image correspond to heart rate. This way, no prior domain knowledge is needed and the system learns the underlying rPPG mechanism from scratch. Work done in [24] implements an image processing pipeline aimed towards extracting a subject's blood volume pulse (BVP) signal, and from that, their pulse rate with a technique called remote photoplethysmography (rPPG). The algorithms are fed a video of a subject, and processed each frame of the video to extract time-indexed RGB vectors. The vectors then go through a pipeline of spectral and statistical analysis algorithms to extract a BVP signal, and from that, their heart rate.

The spectral method implemented in [24] is inspired by [25]and consists of roughly three portions: ROI detection, preprocessing and extraction, and pulse rate calculation. The first of which is aimed to calculate the location of the subject's face to measure the BVP signal. To maintain a robust sequence of measurements on a relatively static portion of the subject's face. After their face has been segmented, each RGB channel in the face image is averaged, resulting in one measurement per channel per frame in the video, producing three signals. As heart rate signals are non-stationary, paper detrends these signals using a smoothness priors approach with a cutoff frequency of 0.33 Hz [26]. After the RGB signals have been detrended and z-normalized, they use Independent Component Analysis (ICA) to decompose them into three independent source signals. ICA separates color variations due to BVP from variations caused by motion, lighting, or other sources. One of the returned components represents the fluctuations in color caused by variations in blood volume. They then filter the signal in the time and frequency domains with a 5 point moving-average filter and a hamming window bandpass filter with cut-off frequencies depending on the user's inputted state. After choosing between resting, recovery, and active; each of which has different cut-off frequencies that incorporate prior estimates of their heart rate. Once the BVP signal is calculated, we use the inter-beat interval estimation implementation described in [27] to estimate the heart rate of the subject.

Multiple datasets for emotion recognition were studied. AMIGOS dataset [28], which stands for A dataset for multimodal research of affect, personality traits, and mood in Individuals and GrOupS. The data were collected from 40 subjects watching videos, with 16 samples each. Biosignals included are ECG, EEG, and GSR. The ECG device used was a Shimmer, at a 256 Hz sampling frequency. The ECG lead configurations used were right arm left leg (RALL), and left arm left leg (LALL). The emotion annotation labels were from a self-assessment and third-person perspectives with a 3D ADM. DREAMER [29], It contains data collected from 23 participants, with 18 samples each. The stimuli used were video clips ranging from 1 to 3 min, with the focus on the ECG and EEG modalities. The ECG device used was a low-cost, wireless, portable, and wearable off-the-shelf device from Shimmer. The sampling rate was 256 Hz, with two-lead and three-lead configurations. Self-annotation of the subjects was conducted using a valence, arousal, and dominance ADM. K-EmoCon [30], contains data collected from 32 subjects in real-time from a naturalistic conversation (paired debates on social issues) to induce emotions. The physiological modalities included are ECG, EEG, BVP, EDA, and skin temperature (SKT). For the ECG signal, a Polar H7 was used, at a 1 Hz sampling rate. The only feature extracted was the HR. This paper claims to be the first publicly available dataset on emotion recognition that has a multi-perspective annotation from selfassessment, second person, and third person. The ADM with valence and arousal scales was implemented. In MANHOB-HCI [31],data were collected from 27 subjects, with 20 samples, using ECG, EEG, GSR, EDA, RSP, and SKT. The ECG device used was a Biosemi Active II, with a three-lead configuration. The sampling rate was 1024 Hz and was downsampled to 256 Hz. Based on the emotional videos watched, the subjects self-reported their affective state with a 3D ADM. SWELL [3]dataset is also known as SWELL knowledge work (SWELL-KW), and it is a new multimodal dataset for research on stress and user modeling. The data were collected from 25 subjects performing tasks such as writing, presenting, reading, and searching to elicit stress. The physiological signals recorded were ECG and SC. The ECG was recorded through a Mobi device (TMSi), with the electrodes placed in a triangular configuration on the chest. The sampling rate was 2048 Hz, with three leads attached. The assessment was conducted by the subjects through labeling two emotional models, which were the ADM and Pos/Neg. WESAD [2] stands for Wearable Stress and Affect Detection. The data were collected from 15 subjects watching video clips and provided with public speaking and mental arithmetic tasks. The biosignals included are ECG, BVP, EDA, EMG, RSP, and temperature (TEMP). The ECG signal was acquired from a RespiBAN Professional using a three-lead configuration. The sampling rate was 700 Hz. The subject self annotated their emotions using a three-class Pos/Neg model. Amusement, neutral, and stress were the classification categories implemented.

3 Proposed method

This paper proposes a solution for wireless emotion recognition which involves wireless extraction of heart rate variability signals and then predicts emotions using an emotion recognizer classification model. Figure 3 gives the architecture of entire system. It consists of 2 major modules : Remote Photoplethysmography

and Emotion Recognizer. Figure 4 describes the proposed method and details of process in each module. For emotion recognizer, we use KNN classifier (comparitive study of logistic regression, Naive bayesian, KNN and SVM is done and KNN outperforms in accuracy of model).

3.1 Baseline Method

Work can be divided in two parts based on modules, baseline model for HRV extraction [24] and a model for emotion classification using time and frequency domain features of HRV signal.

3.1.1 Heart Rate Variability Detection Using Remote Photoplethysmography

Work done in [24] implements an image processing pipeline aimed towards extracting a subject's blood volume pulse (BVP) signal, and from that, their pulse rate with a technique called remote photoplethysmography (rPPG). The algorithms are fed a video of a subject, and processed each frame of the video to extract time-indexed RGB vectors. The vectors then go through a pipeline of spectral and statistical analysis algorithms to extract a BVP signal, and from that, their heart rate.

The spectral method implemented in [24] is inspired by [25]and consists of roughly three portions: ROI detection, preprocessing and extraction, and pulse rate calculation. The first of which is aimed to calculate the location of the subject's face to measure the BVP signal. To maintain a robust sequence of measurements on a relatively static portion of the subject's face. After their face has been segmented, each RGB channel in the face image is averaged, resulting in one measurement per channel per frame in the video, producing three signals. It refers [25] for deriving HRV signals from video. By recording a video of the facial region with a webcam, the red, green, and blue (RGB) color sensors pick up a mixture of the reflected plethysmographic signal along with other sources of fluctuations in light due to artifacts. Given that hemoglobin absorptivity differs across the visible and near-infrared spectral range, each color sensor records a mixture of the original source signals with slightly different weights. Independent Component Analysis (ICA) model assumes that the observed signals are linear mixtures of the underlying source signals. To uncover the independent sources they maximize the non-Gaussianity of each source. The iterative methods are used to maximize or minimize a given cost function that measures non-Gaussianity. As shown in figure 1, OpenCV is used to identify face location in each frame of video and 60% width and full height box is selected on face which is selected ROI. RGB channels are then extracted from ROI for each pixel. Then it detrends these signals using a smoothness priors approach with a cutoff frequency of 0.89 Hz [26].After the RGB signals have been detrended and z-normalized, they use Independent Component Analysis (ICA) to decompose them into three independent source signals. ICA separates color variations due to BVP from variations caused by motion, lighting, or other sources which is highest peak component and selected for further analysis.

Detected high peak signal is then filtered in the time and frequency domains with a 5 point moving-average filter and a hamming window bandpass filter with cut-off frequencies depending on the user's inputted state. After choosing between resting, recovery, and active; each of which has different cut-off frequencies that incorporate prior estimates of their heart rate. Once the BVP signal is calculated, we use the interbeat-interval

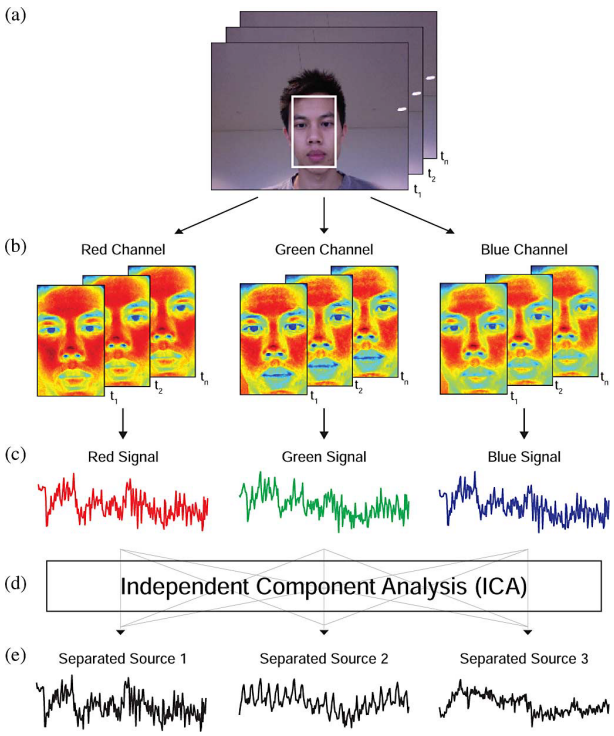

Fig. 1: Recovery of the BVP waveform. (a) Face within the first video frame is automatically detected to locate the ROI. (b) ROI is decomposed into red, green, and blue channels for each frame and spatially averaged to form (c) the raw signals. After the raw signals are detrended and normalized, ICA is applied to separate three independent sources. In this example, the BVP is visible in the second source signal. [24]

estimation implementation described in [27] to estimate the heart rate of the subject. Other heart rate signals are calculated using heartpy python library.

3.1.2 Emotion Recognition from Heart Rate Variability Signals

There are various datasets used for emotion recognition based on physiological signals. But very few pertain to use HRV for emotion recognition. Moreover, for more diverse emotion classes, we have combined 2 datasets SWELL and WESAD for emotion recognition [3] [2]. We shuffled the concatenated dataset to avoid bias towards one dataset. We have renamed emotion classes as numbers in order of ['amusement' 'baseline' 'interruption' 'no stress' 'stress', 'time pressure']. The datasets required Feature scaling hence we use python's standard scaler. The combined dataset faces 2 major issues, imbalance and overlap among classes. We try processing dataset to combat these issues. SMOTE is an oversampling method that synthesizes new plausible examples in the minority class. Tomek Links refers to a method for identifying pairs of nearest neighbors in a dataset that have different classes. Specifically, first the SMOTE method is applied to oversample the minority class to a balanced distribution, then examples in Tomek Links from the majority classes are identified and removed. The combination was shown to provide a reduction in false negatives at the cost of an increase in false positives for a binary classification task [32]. For classification purpose, we use 4 models, logistic regression, KNN, SVM and Naive Bayes and compare their results.

- Logistic Regression: We perform classification of under-sampled , oversampled (SMOTE) data using a liblinear

solver and 'l1' penalty. We use 10-fold cross validation on each run.

- Naive Bayes: We perform classification of undersampled using a Gaussian Naive Bayesian classifier with microaveraging. We use 10-fold cross validation on each run.
- K-nearest neighbour: We perform classification on oversampled (SMOTETomek) data using k=500. Value of k is selected based on lowest error rate obtained during training knn classifier with different values of k as shown in figure 2. We use 10-fold cross validation on each run.
- SVM classifier : We perform classification on oversampled (SMOTETomek) data using RBF and Polynomial kernels. We observe that Radial Basis Function (RBF) kernel gives better classification results than polynomial kernel. We use 10-fold cross validation on each run.

Once it is known that HRV signals are enough for emotion recognition, we used RAVDEES dataset which is video dataset for emotion recognition. Remotephotoplethysmography discussed in previous section was used to obtain HRV and HR features from RAVDEES videos. These obtained signals where used to train the classifier. SmoteTomek oversampling was done on this dataset with feature scaling and 10-fold cross validation to increase the result accuracy.

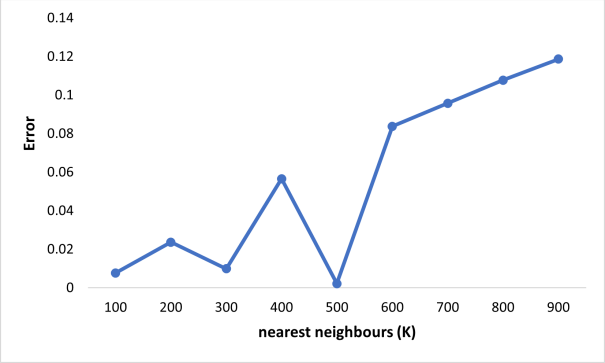

Fig. 2: Error Rate variation for different nearest neighbours values in KNN classification.

3.2 Proposed Model

The Architecture of proposed model is shown in Figure 3. For Wireless emotion recognition, webcam or other video capturing devices can be used. Videos are input for Video to HRV signal extractor. This extractor does spectral analysis of frames in videos and gives time and frequency domain features of heart rate variability. These are then input to a emotion classifier which classifies the subject's emotions belonging to one of the classes in ['amusement' 'baseline' 'interruption' 'no stress' 'stress', 'time pressure']. We use various classifiers and compare their results so as to improve emotion recognizer performance through this paper.

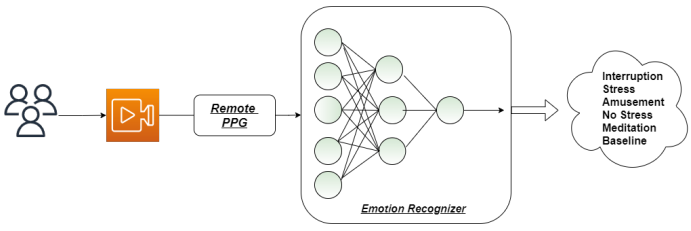

Fig. 3: Wireless Emotion Recognizer Proposed Architecture

3.3 Implementation Details

The Methodology of the proposed model is shown graphically in Figure 3. The Emotion predictor is a trained model using Dataset. We extract Heart Rate Variability time and frequency domain features from videos using rPPG. These features are modified as per the requirements of classifiers and then the classification model is trained. Classification is done using Naive Bayes classifier, KNN classifier, OVR based binary classifiers like Logistic regression, and Support Vector Machines. Classification trained model gets input from the rPPG module which is a video to the HRV extractor. It takes subject's videos and do spectral analysis on them to obtain HRV signals. These videos are divided into image frames. From each frame has face image of a subject time-indexed RGB vectors are extracted. This is carried out by a selection of Region of Interests from the face. To maintain a robust sequence of measurements on a relatively static portion of the subject's face, we used the DLib facial landmark detector to extract the area between the subject's cheeks. These signals are detrended with a cut-off frequency of 0.33Hz and then z normalized. Independent Component Analysis is carried out using scitkit-learn's FastICA. The obtained BVP signal is then filtered to get time and frequency domains in HRV. These features form input to the pre-trained emotion classifier whic then predicts the emotion of the obtained HRV signal. Emotion classifier is trained using processed RAVDEES dataset. The dataset is preprocessed to remove audio from video, remove imbalance in datasets, handle overlapping among classes. Results from various classification techniques are obtained a model is chosen to use as emotion recognizer in our wireless emotion recognizer system.

4 Experimental Setup

For rPPG, all videos were recorded using a Logitech C920 webcam at 30 fps in uncompressed YUYV422 pixel format. PPG signals were collected as ground truth, at a frequency of 60 Hz, via a pulse oxymeter device (CMS50E) attached to the left index finger of the participant.

Classification model training and prediction is carried out on Windows 10 Operating System on 11th Gen Intel(R) Core(TM) i7-11370H @ 3.30GHz 3.00 GHz x64-based processor with 16 GB. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) contains 7356 files (total size: 24.8 GB). The database contains 24 professional actors (12 female, 12 male), vocalizing two lexically-matched statements in a neutral North American accent. Speech includes calm, happy, sad, angry, fearful, surprise, and disgust expressions, and song contains calm, happy, sad, angry, and fearful emotions. Each expression is produced at two levels of emotional intensity (normal, strong), with an additional neutral expression. All conditions are available in three modality formats: Audio-only (16bit, 48kHz .wav), Audio-Video (720p H.264, AAC 48kHz, .mp4), and Video-only (no sound). Note, there are no song files for Actor_18. The datasets are divided into training, testing, and 10-fold cross validation is performed. Entire implementation is done in Python 3.9.7 and all its latest libraries. Swell and Wesad datasets are obtained from [2], [3].

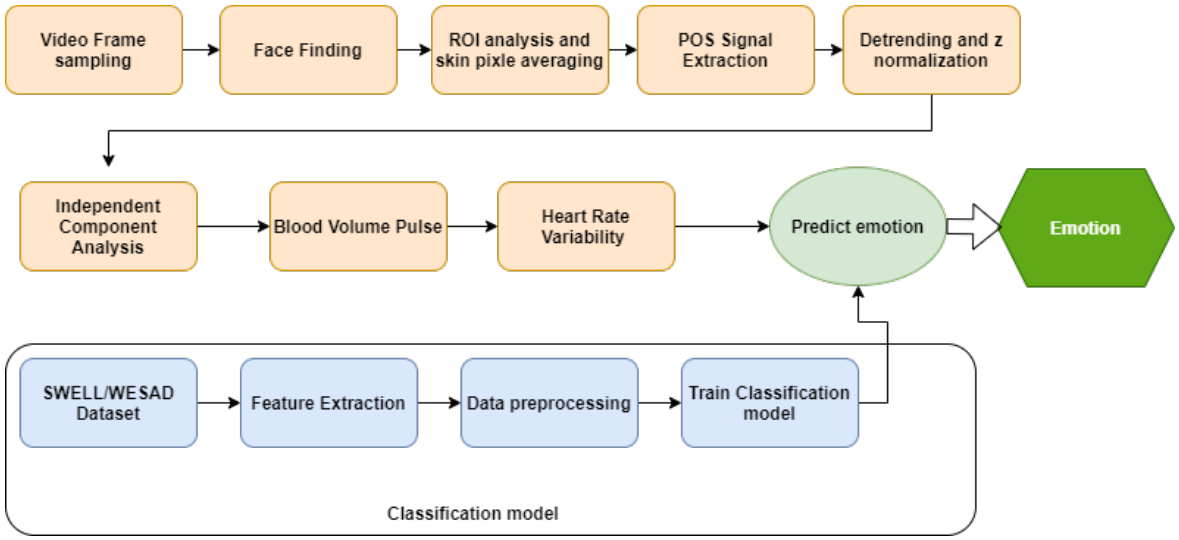

Fig. 4: Proposed Wireless Emotion Recognizer Methodology

4.1 Research Questions

Wireless emotion recognition being a new area of research, there are some challenges to be resolved. Capturing physiological signals remotely gives lower accuracy as compared to wired biosensors. Work is already done on frequency domain signals being used for emotion recognition in [22]. The research area which this paper focuses on is using time-domain signals for emotion recognition. Moreover, this paper aims to improve the accuracy of wireless emotion recognizer such that collective accuracy of remote physiological sensing and emotion classification is increasing. This project works on following questions,

- What features should be considered from HRV for emotion recognition?
- Which model should be selected for classification in imbalanced and overlapped dataset?
- Will the rPPG obtained signals give accuracy higher than or atleast equal to wired or physical methods of emotion recognition?

4.2 Dataset and Preprocessing Techniques

We have used 2 datasets SWELL and WESAD for emotion recognition [3] [2]. Each dataset has classified the emotions into 3 classes. We want to diversify our classification and increase the number of classes. Hence we combine the two datasets.

4.2.1 Datasets

For extracting Heart rate variability signals, we obtain 2560 videos of 24 actors of speech and song from RAVDEES dataset with 8 emotions as 'neutral', 'calm', 'happy', 'sad', 'angry', 'fearful', 'disgust', 'surprised' [1]. Their audios are removed and heart rates were estimated using established rPPG algorithm in [24].
We use the Multivariate, Time-Series dataset, SWELL, and WESAD for training the emotion recognizer [3] [2].SWELL has 391639 records from 25 subjects and WESAD has 106962084 records from 15 subjects under various stress level conditions. This dataset comprises of Heart Rate Variability (HRV) and Electrodermal activity (EDA) features. The SWELL dataset classifies the stress-level emotions as interruption, nostress, timepressure, and WESAD classifies as a baseline, amusement, stress. As a final dataset for training, we combine the two datasets with only HRV features to increase the classification classes as an 'amusement' 'baseline' 'interruption' 'no stress' 'stress', 'time pressure'.

4.2.2 Preprocessing

Figure 6 shows distribution of data where x-axis denotes the classes [0-5] and y-axis denotes the frequency of samples in percentages. We observe the imbalance in dataset where ratio of frequency of samples belonging to class 1 to frequency of samples belonging to class 3 have ratio of 11:89.

Figure 6 shows that the class 1 remains unpredicted and most prediction is caused to class 3, this is because dataset is imbalanced. Since SWELL dataset is smaller than WESAD, frequency of emotions in SWELL are much less than WESAD. It is lowest for emotion 'baseline' which is numbered 1. This causes imbalance in data. Moreover, the records for class 'no-stress' are highest which causes most classifications to be 3.0. Also, since both datasets are whole in their own context and are related to emotion recognition, there is high chance of overlapping classes. Figure 5 shows overlapping of classes in dataset. Due to imbalance, we try balancing using both Undersampling and Oversampling.

- ***Undersampling***: In this type, we equalize the samples of all classes equal to frequency of lowest sample. Here, 'baseline' has lowest samples i.e. 23064. We shuffle the samples and trim all classes to this count and reshuffle again. Reshuffling would lower the possibility of data loss. We observe that without undersampling, logistic regression model was overfitting and class 1 was not fully predicted. Hence we used undersampling and as per the results in table 3, we observe that logistic regression performed better for undersampled data than oversampled or not sampled data.
- ***Oversampling***: We use SMOTE (Synthetic minority oversampling technique) for oversampling. SMOTE generates new samples in between existing data points based on their local density and their borders with the other class. Not only does it perform oversampling, but can subsequently use cleaning techniques (undersampling, more on this shortly) to remove redundancy in the end. [34]. But in

TABLE 1: Heart Rate Variability features description [33]

Feature	Domain	Description
MEAN_ RR	time	Mean of RR intervals
MEDIAN_ RR	time	Median of RR intervals
SDRR	time	Standard deviation of RR intervals
RMSSD	time	Root mean square of successive RR interval differences
SDSD	time	standard deviation of successive RR interval differences
SDRR_RMSSD	time	Standard deviation of RR intervals per Root mean square of successive RR interval difference
HR	time	Heart Rate
pNN25	time	Percentage of successive RR intervals that differ by more than 25ms
pNN50	time	Percentage of successive RR intervals that differ by more than 50ms
SD1	time	Poincare plot standard deviation perpendicular the line of identity
SD2	time	Poincare plot standard deviation along the line of identity

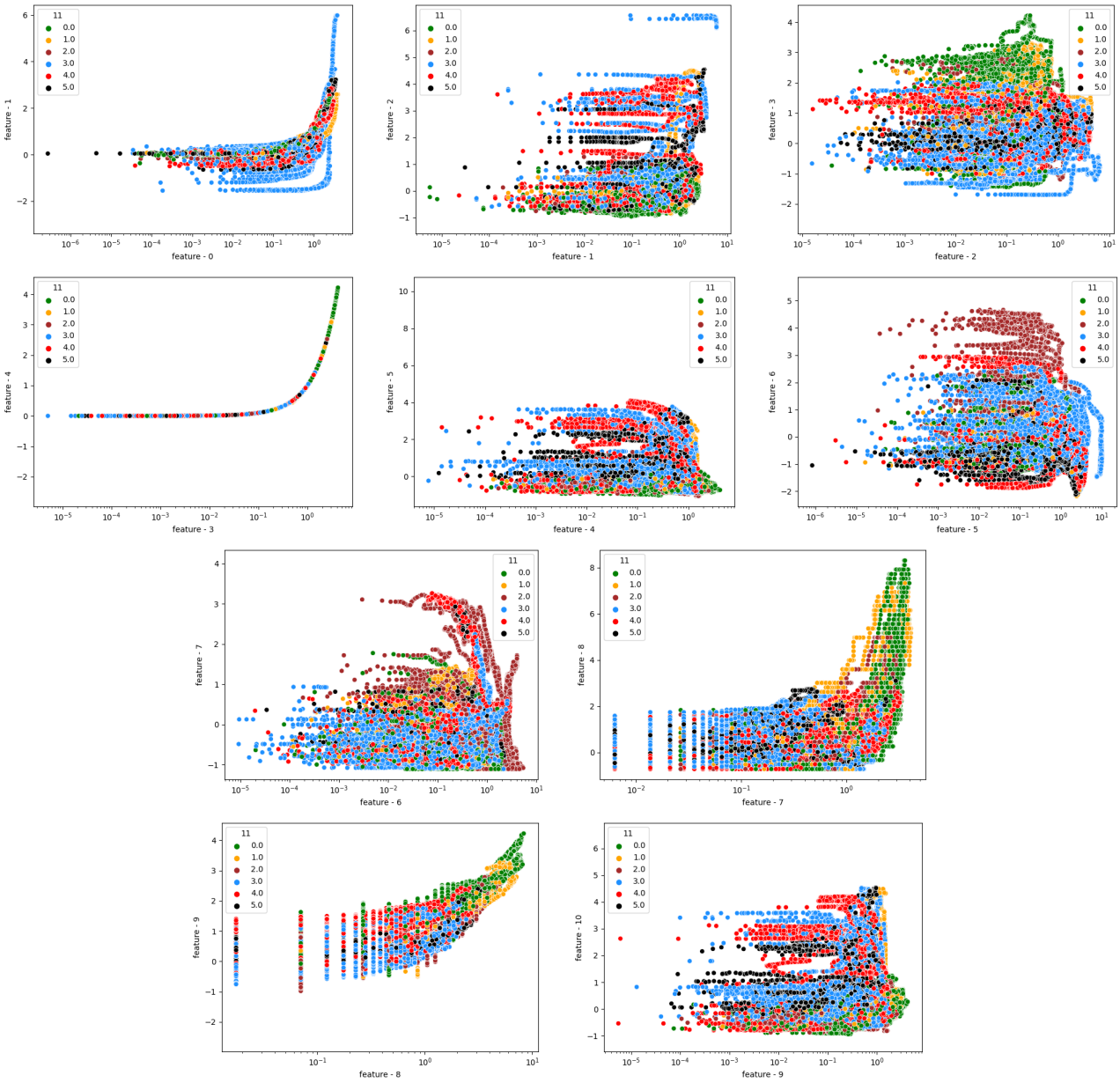

Fig. 5: Visualization of pairwise features of emotion recognition dataset where x and y axis denotes pair of features numbered according to (mean of RR list, median of RR list, sdrr, rmssd, sdsd, hr, pnn25, pnn50, sd1, sd2) and legend shows different class representations. Each subfigure shows high degree of overlap of classes for each pair of feature thus showing a need to handel overlapping in dataset.

case of overlapping classes, SMOTE oversampling fails by inducing more overlap degree. SMOTETomek method discussed in [32] is a solution in this case. It balances the dataset with SMOTE then remove Tomek links from all classes.

4.2.3 Missing Features

Not all features present in datasets [3] [2] which we use for training emotion recognizer were present in the HRV signals extracted from videos using rPPG. [24] provided the ibi , RRlist, and Frequency domain features. Few Time domain features were computed based on RR list. Hence other features like MEDIAN_RR , SDRR, SDRR_RMSSD, HR, pnn25 are computed. [35] [7]

4.2.4 Feature Scaling

Range of features is very diverse in original dataset. This brings the need of feature scaling. We have used standard scaler which standardize features by removing the mean and scaling to unit variance

$$\begin{aligned} \text{MEDIAN\_RR} &= \text{Median (RR\_LIST)} \\ \text{HR} &= \frac{6000}{\text{IBI}} \\ \text{pnn25} &= \frac{x\epsilon\{\text{RR\_DIFF}\}, \forall x|x > 25}{|\text{RR\_DIFF}|} \end{aligned} \tag{1}$$

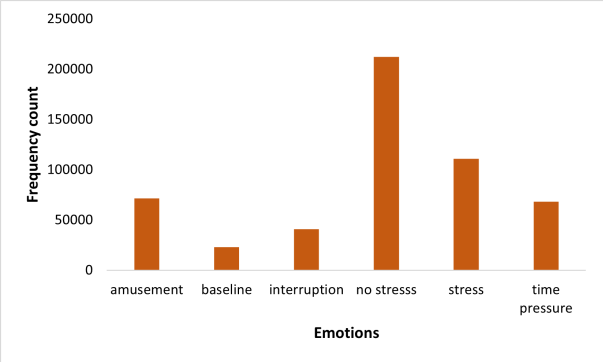

Fig. 6: Data distribution over 6 classes in dataset showing imbalance in dataset where more samples belong to class no stress and thus creating need to handel imbalance.

4.3 Evaluation Metrics

We use classification results using 4 models, Logistic Regression, SVM classifier, KNN classifier and Random Forest. These results are compared based on Accuracy, Recall, Precision and F1 Score.

$$\begin{aligned} \text{Accuracy} &= \frac{\text{Correct Predictions}}{\text{Total Predictions}} \\ \text{Recall} &= \frac{\text{Correct Positive predictions}}{\text{Total Actual positives}} \\ \text{Precision} &= \frac{\text{Correct Positive predictions}}{\text{Total Predicted Positives}} \\ \text{F1 Score} &= \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision + Recall}} \end{aligned} \tag{2}$$

ROC Curves for each classifier is generated and finally, we can assess the performance of the model by the area under the ROC curve (AUC). ROC is a probability curve for different classes. ROC tells us how good the model is for distinguishing the given classes, in terms of the predicted probability. Higher the area under curve, better is the classification.

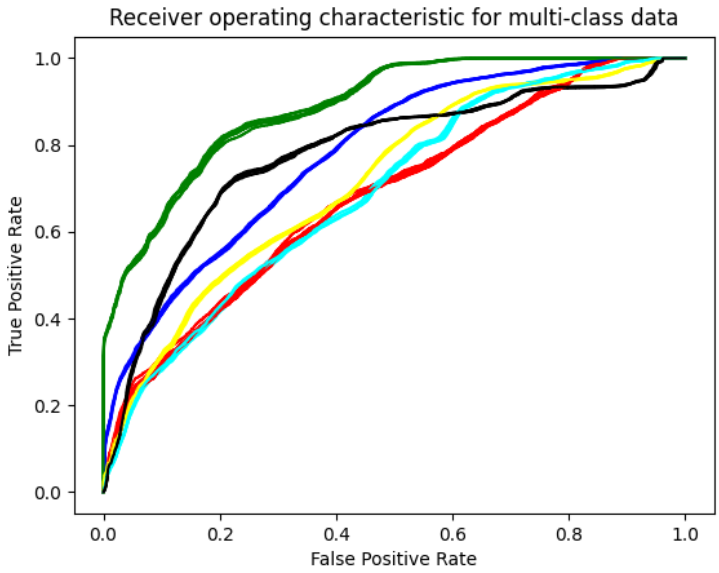

Fig. 7: Figure shows True positive rate vs False positive rate as ROC Curves for each class (0-5) representing AUC for Logistic Regression (one vs all strategy) using SWELL-WESAD dataset. Class 0 vs rest represented by (—), class 1 vs rest represented by (—), class 2 vs rest represented by (—), class 3 vs rest represented by (—), class 4 vs rest represented by (—) and class 5 vs rest represented by (—)

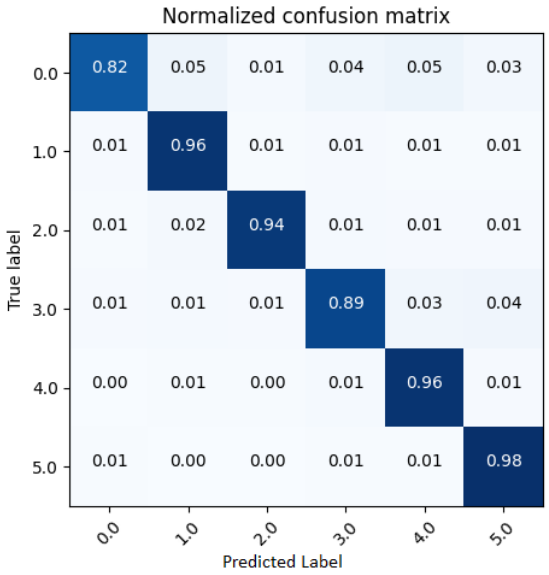

Fig. 8: Figure shows confusion matrix of KNN classifier with k=500 and after using SMOTETomek for data sampling.

5 Results and Discussion

Our proposed model uses KNN classifier with k=500 and smote-tomek oversampling. Figure 2 shows lowest error rate and figure 9 shows highest accuracy obtained by KNN model for nearest neighbours value as 500. Proposed model gives us 93.27% accuracy for SWELL-WESAD dataset which is HRV-emotions dataset. The state of art model considered is a rPPG based pulse rate variability analysis for emotions. Its accuracy being 59.79% using PRV signals obtained using rPPG. We tried rPPG using code in [24] for videos in RAVDEES dataset [1] and later use pretrained model using SWELL-WESAD dataset. A fact to be

TABLE 2: Accuracy and F1-score of Classification Models discussed by various researches with different datasets and different emotion distributions.

Year	Reference	Emotions	Signals	Classification model	Dataset	Testing Accuracy%	F-1 Score%
2019	[12]	6	Facial expressions	DNN	CK+, JAFEE	95.24	-
2020	[13]	4	Facial + EEG	NN, SVM respectively	own	82.75	-
2020	[14]	4	ECG	LS-SVM	own	61.52	-
2021	[15]	2	ECG	Ensemble classifier	AMIGOS	90.2	-
2021	[16]	4	Eye-fixation data	Random Forest	own	80.55	-
2021	[17]	2	PPG-EDA	SVM	own	66.67	-
2019	[20]	2	rPPG-PRV	SVM-RBF	$CAS(ME)^2$	60	-
2019	[20]	4	rPPG-PRV	LBP-TOP,SVM-RBF	$CAS(ME)^2$	40.95	-
2021	[18]	4	Respiration +HRV	Bagged tree classifier	own	63.6	-
2021	[19]	2	PPG-HRV(static)	XGBoost	own	90	-
2021	[19]	2	PPG-HRV(dynamic)	XGBoost	own	80.2	-
**	proposed	6	HRV	Logistic Regression (Undersampling)	SWELL+WESAD	50.34	50.20
**	proposed	6	HRV	Logistic Regression + SMOTE	SWELL+WESAD	41.5	38.94
**	proposed	6	HRV	Naive Bayes (Undersampling)	SWELL+WESAD	35.46	35.43
**	proposed	6	HRV	KNN (k= 500)	SWELL+WESAD	76.56	86.18
**	proposed	6	HRV	KNN + SMOTETomek(k=500)	SWELL+WESAD	93.27	92.91
**	proposed	6	HRV	SVM-RBF + SMOTETomek	SWELL+WESAD	85.36	85.01
**	proposed	6	HRV	SVM-Polynomial + SMOTETomek	SWELL+WESAD	61.71	58.92
**	proposed	8	rPPG+HRV	KNN (k=5) + SMOTETomek	RAVDESS	40.44	40.45

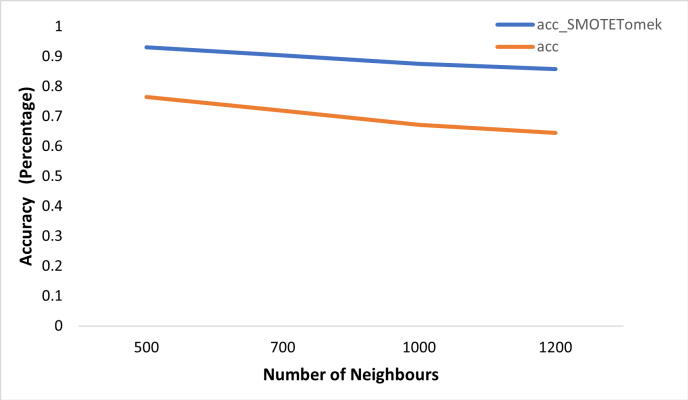

Fig. 9: Figure shows improvement in accuracies after oversampling of SWELL-WESAD dataset using SMOTETomek. Moreover, figure also shows varying accuracy levels for KNN model applied to dataset for different values of nearest-neighbours. This graph helps us to pick value of nearest-neighbour to train KNN model to get highest accuracy.

observed here is rPPG algorithm used by us, proposed in [24] has RMSE = 6.76. With not so accurate HRV predictions, our proposed emotion recognizer gave accuracy of 29.41%. But for accurate HRV features, accuracy obtained is around 93.27%.

We have done evaluation using Logistic Regression, KNN classifier, Naive Bayesian classifier and SVM classifier too. Table 3 show comparative analysis of state-of-art models proposed in previous work and the proposed models for classification.

5.1 Model Evaluation

Logistic Regression and SVM are binary classifiers. For our muli-class dataset, we use OVR strategy for classification in each of these classifiers. Due to imbalance in datasets, we perform sampling. After oversampling using SMOTE, we obtain accuracy of 41.34% and after undersampling accuracy is 50.34%. For imbalanced dataset, one of the appropriate measure is F1 score. We observe that performance of Logistic regression using undersampling is better than performance by SMOTE oversampling. This is because our dataset has overlapping classes and with SMOTE oversampling, more misclassification have been introduced resulting in wrong data samples. Figure 7 shows the ROC-AUC curve using OVR strategy. It is observed that class 2 has highest area under curve which shows ability of model to distinguish between class 2 and others is highest. Each class is distinguished with higher ability since area is more than half. We further use SMOTETomek oversampling method for sampling so as to avoid cons of SMOTE in overlapping classes dataset. With SVM classifier, used two variations, one is Radial Basis Function (RBF) kernel and other is polynomial kernel. Results obtained using RBF kernel (85.36%) are remarkably better than results using logistic regression.

KNN classifier and naive bayesian are a multiclass classifiers. We tried performing classification using KNN with and without oversampling using SMOTETomek. Classification results on imbalanced datasets were lower (76.56%) than one using oversampling technique (93.27%). Figure 9 shows how accuracy varies based on values of nearest neighbours chosen while training KNN model. We observe that for k = 500, we obtain highest values of accuracy. Hence we trained our model using k =500. We show the results in figure 8 for confusion matrix for KNN classifier.

For Naive Bayes classifier, we observed that oversampling provided detoriated results. Hence we performed undersampling and make frequency of all classes equal to class 2 which is lowest in all. We observe accuracy of 35.46% and f-1 score as 35.43%, which is lower than all the proposed models.

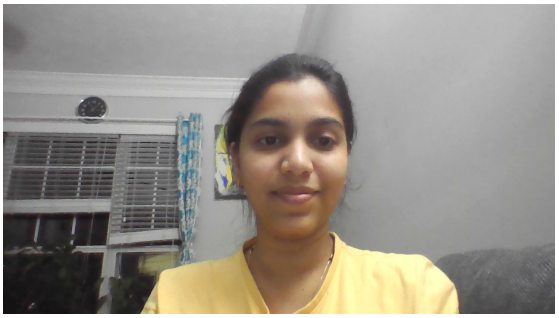

Fig. 10: Figure shows a subject smiling. This photo is a part of video used for testing.

Figure 10 is a snapshot from video which is used to test the wireless emotion recognizer using KNN classifier. We obtain the results as subject emotion is classifier as amusement which is true.

TABLE 3: HRV signals used as features for emotion recognition obtained using rPPG for video with smiling subject in figure 10

Feature	Value
sdsd	126.46
rmssd	193.64
pnn50	0.75
sd1	136.89
sd2	209.81
MEAN_RR	636.86
MEDIAN_RR	620.098
SDRR	167.22
SDRR_RMSSD	0.86
HR	9.42
pnn25	1.0

We also tested this trained model for videos in RAVDESS dataset from paper [1]. It has videos of people with who are singing and speaking 8 emotions (neutral,calm,happy,sad, angry, fearful, disgust, surprised). We removed audio from these videos using [36] since we observed that audio affects HRV detection. We mapped the emotions in RAVDESS dataset to emotions SWELL-WESAD dataset and observed 40.44% accuracy using same KNN model trained on SWELL-WESAD dataset.

6 Conclusion and Future Scope

In this paper, we have proposed a model for wireless emotion recognition using remote photoplethysmography. We derive HRV signals in both time and frequency domain from videos and classify these signals to recognize emotions. For training emotion classifier, we have evaluated using logistic regressor, a KNN classifier tuned parameter k=500, SVM based on RBF kernel and Naive bayesian classifier. Due to imbalance in dataset we have sampled the data. Moreover, due to overlap in dataset, we use SMOTETomek.

Considering the rPPG gives accurate HRV extraction for which we recognize emotions, we tested our KNN model with testing set in SWELL-WESAD dataset. As per the results, we obtain 93.27% accuracy. We also fetched results using logistic regression with undersampled dataset to provide better results as 50.34% accuracy and 50.20% f1 score, SVM using RBF gives 85.36% accuracy. Lowest accuracy among proposed models is obtained by NB model. We also tested the same emotion recognizer model on emotion based videos of RAVDEES after processing them using rPPG for which we obtained accuracy of 40.44%. We observe that using HRV signals gave better for emotion recognition than using PRV signals as in state-of-art model.

We have tested our model using videos in which the subject look into camera. We plan to improve this and test using the heart rate variability signals obtained from subjects who are not directly looking in camera or they are busy doing their own task such that camera can directly take their video for emotion recognition. This will improve the scope of use of this technology. As part of future work, we also plan to increase the diversity of emotion classes and include arousal-valence scale to it.

References

[1] S. R. Livingstone and F. A. Russo, "The ryerson audio-visual database of emotional speech and song (ravdess): A dynamic, multimodal set of facial and vocal expressions in north american english," *PloS one*, vol. 13, no. 5, p. e0196391, 2018.

[2] S. Koldijk, M. Sappelli, S. Verberne, M. A. Neerincx, and W. Kraaij, "The swell knowledge work dataset for stress and user modeling research," in *Proceedings of the 16th international conference on multimodal interaction*, 2014, pp. 291–298.

[3] "Kaggle dataset." [Online]. Available: https://www.kaggle.com/qiriro/stress

[4] T. L. Nwe, S. W. Foo, and L. C. De Silva, "Speech emotion recognition using hidden markov models," *Speech communication*, vol. 41, no. 4, pp. 603–623, 2003.

[5] Y. St-Pierre, C. V. Themsche, and P.-O. Estève, "Emerging features in the regulation of mmp-9 gene expression for the development of novel molecular targets and therapeutic strategies," *Current Drug Targets-Inflammation & Allergy*, vol. 2, no. 3, pp. 206–215, 2003.

[6] S. Wioleta, "Using physiological signals for emotion recognition," in *2013 6th International Conference on Human System Interactions (HSI)*. IEEE, 2013, pp. 556–561.

[7] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, p. 258, 2017.

[8] K.-H. Choi, J. Kim, O. S. Kwon, M. J. Kim, Y. H. Ryu, and J.-E. Park, "Is heart rate variability (hrv) an adequate tool for evaluating human emotions?–a focus on the use of the international affective picture system (iaps)," *Psychiatry research*, vol. 251, pp. 192–196, 2017.

[9] D. S. Quintana, A. J. Guastella, T. Outhred, I. B. Hickie, and A. H. Kemp, "Heart rate variability is associated with emotion recognition: direct evidence for a relationship between the autonomic nervous system and social cognition," *International journal of psychophysiology*, vol. 86, no. 2, pp. 168–172, 2012.

[10] G. Chanel, J. J. Kierkels, M. Soleymani, and T. Pun, "Short-term emotion assessment in a recall paradigm," *International Journal of Human-Computer Studies*, vol. 67, no. 8, pp. 607–627, 2009.

[11] B. L. Sheaffer, J. A. Golden, and P. Averett, "Facial expression recognition deficits and faulty learning: Implications for theoretical models and clinical applications." *International Journal of Behavioral Consultation and Therapy*, vol. 5, no. 1, p. 31, 2009.

[12] D. K. Jain, P. Shamsolmoali, and P. Sehdev, "Extended deep neural network for facial emotion recognition," *Pattern Recognition Letters*, vol. 120, pp. 69–74, 2019.

[13] Y. Huang, J. Yang, P. Liao, and J. Pan, "Fusion of facial expressions and eeg for multimodal emotion recognition," *Computational intelligence and neuroscience*, vol. 2017, 2017.

[14] Y.-L. Hsu, J.-S. Wang, W.-C. Chiang, and C.-H. Hung, "Automatic ecg-based emotion recognition in music listening," *IEEE Transactions on Affective Computing*, vol. 11, no. 1, pp. 85–99, 2017.

[15] A. Sepúlveda, F. Castillo, C. Palma, and M. Rodriguez-Fernandez, "Emotion recognition from ecg signals using wavelet scattering and machine learning," *Applied Sciences*, vol. 11, no. 11, p. 4945, 2021.

[16] L. J. Zheng, J. Mountstephens, and J. Teo, "A comparative investigation of eye fixation-based 4-class emotion recognition in virtual reality using machine learning," in *2021 11th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*. IEEE, 2021, pp. 19–22.

[17] G. Cosoli, A. Poli, L. Scalise, and S. Spinsante, "Heart rate variability analysis with wearable devices: Influence of artifact correction method on classification accuracy for emotion recognition," in *2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE, 2021, pp. 1–6.

[18] L. Zhang, C.-H. Fu, H. Hong, B. Xue, X. Gu, X. Zhu, and C. Li, "Non-contact dual-modality emotion recognition system by cw radar and rgb camera," *IEEE Sensors Journal*, vol. 21, no. 20, pp. 23 198–23 212, 2021.

[19] T. Tagnithammou, É. Monacelli, A. Ferszterowski, and L. Trénoras, "Emotional state detection on mobility vehicle using camera: Feasibility and evaluation study," *Biomedical Signal Processing and Control*, vol. 66, p. 102419, 2021.

[20] R. M. Sabour, Y. Benezeth, F. Marzani, K. Nakamura, R. Gomez, and F. Yang, "Emotional state classification using pulse rate variability," in *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*. IEEE, 2019, pp. 86–90.

[21] J. Gideon and S. Stent, "The way to my heart is through contrastive learning: Remote photoplethysmography from unlabelled video," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3995–4004.

[22] B. Kossack, E. Wisotzky, A. Hilsmann, and P. Eisert, "Automatic region-based heart rate measurement using remote photoplethysmography," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2755–2759.

[23] R. Song, H. Chen, J. Cheng, C. Li, Y. Liu, and X. Chen, "Pulsegan: Learning to generate realistic pulse waveforms in remote photoplethys-

mography," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1373–1384, 2021.

[24] H. P. David Hass, Spencer Mullinix, "Heart rate detection using remote photoplethysmography." [Online]. Available: https://github.com/mullisd1/CV_Heartrate

[25] 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, 2010.

[26] M. P. Tarvainen, P. O. Ranta-Aho, and P. A. Karjalainen, "An advanced detrending method with application to hrv analysis," *IEEE transactions on biomedical engineering*, vol. 49, no. 2, pp. 172–175, 2002.

[27] P. van Gent, H. Farah, N. van Nes, and B. van Arem, "Analysing noisy driver physiology real-time using off-the-shelf sensors: Heart rate analysis software from the taking the fast lane project," *Journal of Open Research Software*, vol. 7, no. 1, 2019.

[28] J. A. M. Correa, M. K. Abadi, N. Sebe, and I. Patras, "Amigos: A dataset for affect, personality and mood research on individuals and groups," *IEEE Transactions on Affective Computing*, 2018.

[29] S. Katsigiannis and N. Ramzan, "Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices," *IEEE journal of biomedical and health informatics*, vol. 22, no. 1, pp. 98–107, 2017.

[30] C. Y. Park, N. Cha, S. Kang, A. Kim, A. H. Khandoker, L. Hadjileontiadis, A. Oh, Y. Jeong, and U. Lee, "K-emocon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations," *Scientific Data*, vol. 7, no. 1, pp. 1–16, 2020.

[31] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 42–55, 2011.

[32] J. Brownlee, "How to combine oversampling and undersampling for imbalanced classification." [Online]. Available: https://machinelearningmastery.com/combine-oversampling-and-undersampling-for-imbalanced-classification/

[33] J. P. G. Fred Shaffer, "An overview of heart rate variability metrics and norms." [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5624990/

[34] M. Stewart, "Behavioral research blog." [Online]. Available: https://towardsdatascience.com/guide-to-classification-on-imbalanced-datasets-d6653aa5fa23

[35] "Psych data." [Online]. Available: http://www.psylab.com/html/default_heartrat.htm

[36] B. Wayne, "How to remove original audio from the video file ?" [Online]. Available: https://github.com/Zulko/moviepy/issues/504