

MULTIMODAL & CROSS-DEVICE EMOTION UNDERSTANDING  
WITH PRIVACY-PRESERVATION  
FOR AFFECTIVE USER EXPERIENCE BASED APPLICATIONS

by

BARATH RAJ KANDUR RAJA,  
MS (Data Science), PG Diploma (Data Science), BE (Computer Science & Engineering)

DISSERTATION

Presented to the Swiss School of Business and Management Geneva

In Partial Fulfillment

Of the Requirements

For the Degree

GLOBAL DOCTOR OF BUSINESS ADMINISTRATION

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

DECEMBER, 2023

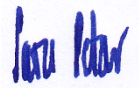
MULTIMODAL & CROSS-DEVICE EMOTION UNDERSTANDING  
WITH PRIVACY-PRESERVATION  
FOR AFFECTIVE USER EXPERIENCE BASED APPLICATIONS

by

BARATH RAJ KANDUR RAJA

APPROVED BY

Prof.dr.sc. Saša Petar, Ph.D., Chair



RECEIVED/APPROVED BY:

---

<Associate Dean's Name, Degree>, Associate Dean

## **Acknowledgements**

I would like to express my gratitude to all the people that supported my research:

- To my **Wife Chandni Rajan**, for staying with me through thick and thin, for encouraging and pursuing my Doctorate,
- To **Prof. Dr. Mario Silic** for being very supportive, encouraging, and for guiding in completing this research thesis,
- To the **Colleagues in Samsung R&D Institute India Bangalore**, for providing me innovative and excellent work environment,
- To **Swiss School of Business and Management (SSBM), Geneva**, for providing this opportunity,
- And finally, to all the researchers, innovators and open-source contributors, for sharing their resources and knowledge!

ABSTRACT

MULTIMODAL & CROSS-DEVICE EMOTION UNDERSTANDING  
WITH PRIVACY-PRESERVATION  
FOR AFFECTIVE USER EXPERIENCE BASED APPLICATIONS

BARATH RAJ KANDUR RAJA  
2023

Dissertation Chair: <Chair's Name>  
Co-Chair: <If applicable. Co-Chair's Name>

Emotion identification is a complex research area that can enable unique multi-device experiences. Smartphones, the dominant mode of communication, can aid in emotion prediction. However, there is a lack of datasets with precise ground truth labels based on user smartphone behavior due to challenges associated with dataset annotation. Present annotation techniques rely either on self-reporting or recording on desktop applications, which is less natural. In this research, these issues are addressed by devising a user-centric approach to collect and annotate user data in a non-intrusive way on smartphones. The insights are derived from the annotated data comprising behavior, emotion, and personality. The data consists of categorical features that do not include personally identifiable information, thus preserving user privacy. The annotated data is validated by an emotion prediction model using the Random Forest classifier, achieving an accuracy score of 67.73%. Further, an accuracy of 77.95% is achieved on sentiment prediction (positive, negative, and neutral) using the Support Vector Machine (SVM) classifier.

## TABLE OF CONTENTS

List of Tables .....	viii
List of Figures.....	x
<b>CHAPTER I: INTRODUCTION .....</b>	<b>1</b>
1.1 Introduction.....	1
<b>CHAPTER II: LITERATURE REVIEW .....</b>	<b>4</b>
2.1 Literature Review .....	4
2.2 Literature Review: Emotion in different Modalities.....	4
2.3 Literature Review Details: Emotion Context from various Modalities .....	8
2.3.1 Textual Emotion Recognition – Directed Acyclic Graph .....	8
2.3.2 Audio Emotion Recognition – Convolutional Neural Network .....	9
2.3.3 Audio Emotion Recognition – Empirical Mode Decomposition.....	10
2.3.4 Image Emotion Recognition using Attentional Convolutional Network .....	11
2.3.5 Video Emotion Recognition – Video-Audio-Text Transformer .....	13
2.3.6 Multimodal Emotion Recognition – Multimodal Transformers.....	15
2.3.7 Video Emotion Recognition – An identity-free video dataset.....	16
2.3.8 Sensor based Emotion Recognition .....	18
2.3.9 Emotion Prediction from Categorical Data .....	19
2.3.10 Rule based ML: Representation Learning .....	20
2.3.11 Image Emotion Recognition – Context based Emotion Recognition.....	21
2.3.12 Self Supervised Learning (SSL) .....	24
2.4 Literature Review Summary: Data collection .....	26
2.5 Literature Survey Details: Emotion from Categorical Data .....	28
2.5.1 MoodScope: Building a Mood Sensor from Smartphone Usage Patterns .....	28
Source: LiKamWa et al. (2013).....	28
2.5.2 iSelf: Towards Cold-start Emotion Labeling using Transfer Learning with Smartphones.....	29
2.5.3 Daily Stress Recognition from Mobile Phone Data, Weather Conditions and Individual Traits .....	30
2.5.4 Does Smartphone Use drive our emotions or vice versa? A Casual Analysis.....	31

2.5.5 Predicting Personality from patterns of behavior collected with smartphones.....	32
2.5.6 Predicting Negative Emotions based on Mobile Phone Usage Patterns: An Explanatory Study.....	33
2.5.7 Emotion Recognition using Mobile phones .....	34
2.6 Summary .....	36
<b>CHAPTER III: RESEARCH MOTIVATION &amp; OBJECTIVES.....</b>	<b>37</b>
3.1 Opportunity Areas .....	37
3.2 Discussion.....	38
3.3 Research Questions .....	38
3.4 Selected Problem Statements & Objectives .....	39
3.5 Data Collection Objectives.....	40
3.6 Categorical Features Selection .....	41
3.7 Data Collection: Risk and Mitigation.....	49
<b>CHAPTER IV: METHODOLOGY.....</b>	<b>50</b>
4.1 Designing Emotion Data Collection Experience .....	50
4.2 System Design and Data Collection .....	54
4.3 Emotional Data Analysis.....	56
4.3.1 Research Context.....	57
4.3.2 Data Visualization and Emotional Insights .....	58
4.4 Feature Engineering .....	62
4.5 Emotion Model – Model Experiments & Preliminary Results .....	65
4.5.1 Release stats – 15 Mar 2023 (Version 4.0, Model Version 1).....	66
4.6 Personality Model – Model Experiments & Preliminary Results.....	68
4.6.1 Personality - Release stats (Version 4.0 Model Version 1).....	69
4.6.2 Personality - User Feedback Results: Version 4.0 Model Version 1 (Till 16 Jun 2023).....	69
4.6.3 Tabnet Architecture .....	70
4.7 Graphical Convolution Network (GCN) Experiment.....	71
4.8 Current Machine Learning Model Architecture .....	72
4.9 Model Upgrades .....	76
4.9.1 Release stats – 16 May 2023 (Version 4.0, Model Version 1).....	76
4.9.2 Release stats – 21 Jun 2023 (Version 4.1, Model Version 2).....	77
<b>CHAPTER V: EMOTION RECOGNITION: RESULTS AND ANALYSIS.....</b>	<b>79</b>
5.1 Emotion Prediction: Result & Analysis.....	79
<b>CHAPTER VI: DISCUSSION .....</b>	<b>86</b>
6.1 Research Question One .....	86

6.2 Research Question Two .....	86
6.3 Research Question Three .....	87
6.4 Research Question Four .....	89
6.5 Research Question Five .....	90
6.6 Research Question Six .....	92
6.7 Summary of Findings.....	93
 CHAPTER VII: SUMMARY AND RECOMMENDATIONS .....	 94
7.1 Conclusion.....	94
7.2 Recommendations for Future Research .....	94
 APPENDIX A IBIS 2023 CONFERENCE .....	 96
APPENDIX B SURVEY: DEMOGRAPHICS & PERSONALITY TRAITS .....	97
APPENDIX C INFORMED CONSENT: PERMISSIONS .....	98
APPENDIX D INFORMED CONSENT: PRIVACY POLICY .....	99
REFERENCES .....	101

## LIST OF TABLES

Table 2.1 Emotion from different Domains / Modalities .....	7
Table 2.2 Results Directed Acyclic Graph Network for Conversational Emotion Recognition.....	9
Table 2.3 Emotion Classification Accuracy .....	12
Table 2.4 Twenty-six Discrete Emotion Categories.....	22
Table 2.5 Methods for Data Collection.....	28
Table 3.1 List of Categorical Features.....	41
Table 4.1 Application Categories .....	63
Table 4.2 Emotion Model Results (Precision, Recall, F1-Score): Mar 2023 .....	66
Table 4.3 Emotion Model - Confusion Matrix: Mar 2023.....	67
Table 4.4 Personality Model Results (Precision, Recall, F1-Score): Mar 2023 .....	69
Table 4.5 Personality Model Results (Accuracy): Jun 2023.....	70
Table 4.6 Personality Model Results (Accuracy) – Tabnet Architecture: Jun 2023 .....	70
Table 4.7 GCN Results for Emotion Prediction .....	72
Table 4.8 Emotion Prediction: Without Demography and Personality .....	73
Table 4.9 Emotion Prediction Improvements: Version 4.0 and Version 4.1 .....	74
Table 4.10 Emotion Model Results (Precision, Recall, F1-Score): Mar 2023 .....	74
Table 4.11 Emotion Model Results (Precision, Recall, F1-Score): May 2023.....	76
Table 4.12 Emotion Model - Confusion Matrix: May 2023.....	77
Table 4.13 Emotion Model Results (Precision, Recall, F1-Score): Jun 2023 .....	77
Table 4.14 Emotion Model - Confusion Matrix: Jun 2023 .....	78
Table 5.1 Comparative study between various smartphone-based annotation applications .....	79
Table 5.2 Performance of different classifiers on Emotion and Sentiment prediction tasks.....	81
Table 5.3 Performance results for Emotion and Sentiment prediction tasks .....	82
Table 5.4 Ablation Study for Emotion Task.....	82
Table B.1 Survey: Demographics.....	97
Table B.2 Survey: Demographics Filled .....	97
Table B.3 Survey: Personality .....	97



Table C.1 Informed Consent: App Permissions (No Personally Identifiable Information).....	98
Table D.1 Informed Consent: Onboarding screen.....	99
Table D.2 Informed Consent: Privacy Policy – Screen 1 .....	99
Table D.3 Informed Consent: Privacy Policy – Screen 2 .....	99
Table D.4 Informed Consent: Privacy Policy – Screen 3 .....	100
Table D.5 Informed Consent: Privacy Policy - Screen 4.....	100

## LIST OF FIGURES

Figure 1.1 Emotion Models: Introduction .....	2
Figure 2.1 Conversational Emotion Recognition with Directed Acyclic Graph Neural Network.....	8
Figure 2.2 Speech Emotion Recognition with Deep CNN .....	9
Figure 2.3 Audio Emotion Detection with Empirical Mode Decomposition on TESS Dataset .....	10
Figure 2.4 Image Emotion Recognition - Architecture Pipeline .....	11
Figure 2.5 Video / Multimodal Transformer Architecture for Human Emotion Classification .....	13
Figure 2.6 Multimodal Emotion Recognition – Cross-Modal Multi-head Attention .....	15
Figure 2.7 Multimodal Emotion Recognition: Cross-Modal Transformer Architecture & Attention .....	16
Figure 2.8 Video Dataset – identity free micro-gestures (Cover-face, fold-arms, cross-fingers) .....	16
Figure 2.9 Video Emotion Recognition – Unsupervised Encoder-Decoder Network .....	17
Figure 2.10 Sensors based Emotion Prediction .....	19
Figure 2.11 Categorical Emotion Prediction - Graph Convolutional Network .....	19
Figure 2.12 Rule-based Representation Learning – Example .....	21
Figure 2.13 Context based Emotion Recognition.....	23
Figure 2.14 MoodScope: Use case and Engine .....	28
Figure 2.15 iSelf: System Architecture and Feature Vector.....	29
Figure 2.16 Daily Stress Recognition: Basic Features and Bluetooth Proximity Features.....	30
Figure 2.17 Smartphone Application Launch & Usage and Emotions .....	31
Figure 2.18 Big 5 Personality Prediction Model .....	32
Figure 2.19 Predicting Negative Emotion: Emotion scales and Feature selection .....	33
Figure 2.20 Emotion Recognition: Data Capture, Prediction and Results.....	35
Figure 2.21 Emotion Recognition: Decision Tree (J48), System Architecture and Experience .....	35
Figure 4.1 Main Screens of the Reflektion Application .....	51

Figure 4.2 Periodic notifications for data collection & Widget nudge for Data annotation .....	52
Figure 4.3 System Design.....	56
Figure 4.4 Data Processing Pipeline .....	57
Figure 4.5 Ground Truth Labels Tagging & Demographic Data Distribution .....	57
Figure 4.6 Emotion Distribution labelled by the Users .....	58
Figure 4.7 Distribution of application categories (Launch count & Usage duration).....	59
Figure 4.8 Emotion Distribution in different category of applications (App Usage vs Emotion).....	60
Figure 4.9 Emotion Distribution among various smart phone features .....	61
Figure 4.10 User Emotion Dynamics .....	61
Figure 4.11 Overview of Model Prediction Pipeline, with Novel smartphone data.....	62
Figure 4.12 Graphical Convolution Network (GCN) Architecture: Experiment .....	71
Figure 4.13 Model Architecture (Emotion Prediction from User Data, Demographics, Personality).....	73
Figure 5.1 t-SNE plot for emotion and sentiment prediction tasks using categorical features .....	84
Figure 5.2 Performance of Random Forest classifier on Emotion Prediction task.....	85
Figure A.1 IBIS 2023 Conference: Presentation Certificate .....	96

## CHAPTER I: INTRODUCTION

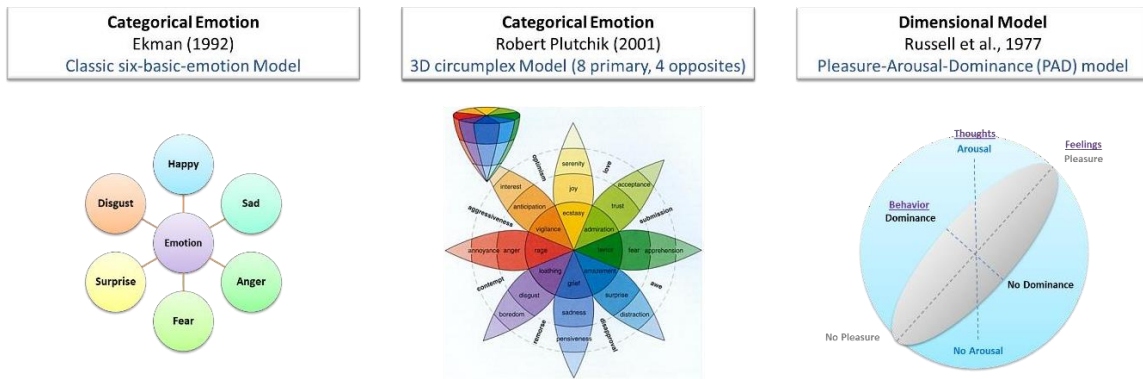
### 1.1 Introduction

Since the 17th century, the word “**emotion**” has been used in English as a translation of the French word “*émotion*”. The term began to be used considerably more frequently in 18th century English, often to refer to mental experiences. In the next century, it ultimately developed into a theoretical term due to the significant influence of two Scottish philosopher-physicians, Thomas Brown and Charles Bell (Dixon, 2012). Over the period, multiple researchers tried to develop models for understanding emotion. Broadly, two types of emotion models are considered: categorical and dimensional. Categorical emotion models divide emotions into distinct categories. Ekman’s (1992) classic six -basic-emotion model comprising of emotions: happy, sad, anger, fear, surprise, and disgust is an example of the same. Robert Plutchik (2001) proposed another prominent categorical emotion model that described emotions as a combination of eight basic emotions. Compared to Categorical Models, Dimensional Models express emotions using a multi-dimensional space. The Pleasure-Arousal-Dominance model (Russell et al., 1977) is one such dimensional model that argues that three dimensions of pleasure-displeasure, arousal-nonarousal, and dominance-submissiveness are sufficient to describe a large variety of emotional states.

Contrary to emotion, which is short-lived, **personality traits** are the cognitive, behavioral, emotional, and motivational characteristics of an individual that are more consistent across situations (McAdams, 2015). Each person varies along a spectrum on any

given trait (e.g., a person can be open-minded, close-minded, or in between). Over the years, many theories have suggested the number of traits that can describe any individual's personality. There are different taxonomies discussed in the literature (Oliver et al., 1999). **In this proposed research, we have focused on the most widely used big five human personality traits: extraversion, openness, conscientiousness, agreeableness, and neuroticism (McCrae et al., 2008).**

*Figure 1.1*  
*Emotion Models: Introduction*



Source: Adapted from Ekman (1992), Plutchik (2001), Adapted from Bakker et al. (2014)

Few attempts have been made to establish a **relationship between personality and emotion**, both equally important to understand the user. The Personality Emotion Model (PEM) workflow enables bi-directional personality-emotion mappings (Donovan et al., 2021). Sadeghian et al. (2021) shows that for emotional profiling or user behavior understanding, the personality can help in boosting the emotional model's accuracy. **The proposed research work attempts to build over this understanding and provide a more practical and non-intrusive way to understand this relation.**

**To study this relation, the user behavior on the smartphone needs to be observed in this research.** Smartphone has become one of the primary companions of the user. It has become an indispensable part of daily life, be it for business or entertainment purposes. People are spending a third of their waking time on smartphones as per app monitoring firm App Annie (data.ai, 2022), which is close to 4.8 hours. Thus, it is essential to understand the factors that can influence a user's emotions in a non-intrusive manner. It could be either way. A user under a particular emotional state tends to use the smartphone in a certain way. It could also be using a smartphone in a certain way can usher specific emotions in the user.

Another important aspect while researching the relation is to **maintain user privacy**. Privacy of personally identifiable information (PII) and other user data is one of the significant concerns that permeate recent technological developments and associated regulations (Pelteret et al., 2016). **Most of the previous works involving the task of emotion prediction from Smartphone usage utilize one or more PII, which may not be desirable to the user.**

## CHAPTER II: LITERATURE REVIEW

### 2.1 Literature Review

Conventionally, Emotion recognition includes text as an input parameter. PS et al. (2017) provided an overview of research on computational emotion models that explains comprehensive survey prior work performed in this field. In summary, they have presented two types of models for conveying emotions. Firstly, the categorical model, where the emotions are labelled. Secondly, the dimensional model, where the representation is based on multi-dimensional scaling which is a quantitative approach. It is, in the state of the art, proved that the Categorical Emotion model is best suited for Human Emotion Recognition.

In this literature review, the prior work and the state-of-the-art (SOTA) models for different modalities (modalities include text, audio, image, video, etc.) is performed. In addition, public data set available for each of the modalities are explored.

### 2.2 Literature Review: Emotion in different Modalities

Felbo et al (2017) experimented a deep learning based SOTA architecture called ‘DeepMoji’ model, that uses millions of tweets (**text**) with emojis, to detect sentiment, emotion and sarcasm. This is based on emotion understanding of a single sentence. Later, model architectures with Directed Acyclic Graph based Network (DAG-ERC) is experimented, to understand intrinsic structure of emotions in **conversations** (Shen et al., 2021). In DAG-ERC, each sentence is constructed as a graph node, whose features are extracted by Transformer-based Language Models. Further, in this paper, performance of datasets like *IEMOCAP*, *MELD*, *DailyDialog*, *EmoryNLP* are evaluated. The latest models include DialogueRNN (Majumder et al., 2019) and DialogueGCN (Ghosal et al., 2019),

which outperformed various SOTA contextual emotion classifiers for emotion recognition in conversations.

For understanding emotions in **Audio**, Issa et al (2020) explored speech emotion recognition with deep convolutional neural networks, on *RAVDESS* dataset. Here, five-fold cross-validation is used for training the model, in which data set is randomly divided into five groups, to make the classification speaker-independent, with conventional 80% training dataset and 20% testing dataset. This schema is performed five times and the average classification accuracy on test sets is computed. Further, Krishnan et al (2021) evaluated Empirical Mode Decomposition (EMD) and non-linear features as input for SOTA machine learning classifiers, that are trained on entropy features. Several classifiers such as SVM, KNN, LDA, Naïve Bayes, Gradient Boosting, Random Forest are compared against Mean Balanced Accuracy (MBA), and it was proved that, among all, LDA performed best, on *TESS* dataset.

Attentional Convolutional Network (Minaee et al., 2021) is experimented for recognizing deep emotions using **Facial** expressions. This is built on the observation that not all the parts of the face are important for detecting specific emotion. Thus, it utilizes feature extraction with spatial transformers (localization network) to focus on important facial regions. The network also utilizes few layers for faster inference speed and more suitable for real-time applications. Some of **Image** datasets found for emotion recognition are *FER2013*, *CK+*, *JAFFE*, and *FERG*.

Yue-Hei Ng et al. (2015) proposed Deep Networks for **Video** classification problems. Two methods of video classification using CNNs are proposed. The first method



explores various convolutional temporal feature based pooling architectures, examining the various design choices which need to be made, when adapting CNN for the task. The second proposed method explicitly models the video as an ordered sequence of frames using LSTM networks. Carreira et al. (2017) delivered a new model and *KINETICS* dataset for action recognition in videos. I3D architecture is proposed in this paper, which builds upon SOTA Image classification architectures, but inflates the filters and pooling kernels into 3D, leading to a very deep, naturally spatiotemporal classifiers. Since 3D convolutions are expensive, Xie et al. (2018) tried to achieve a speed-accuracy tradeoff by converting some convolutions to 2D and by introducing S3D (separable spatial and temporal convolutions).

Finally, Akbari et al. (2021) proposed Video-Audio-Text Transformer (VATT), for learning **multimodal** representations from unlabeled data using convolution-free Transformer architectures. The backbone network comprises of a regular transform architecture (BERT), without using any CNN layers, which makes it more suitable multimodal data. Since unsupervised learning (contrastive loss) is utilized, we can benefit from the large dataset of unlabeled videos (ex: YouTube videos). Such a pretrained network which has learnt the common representations from video and audio data, can be used for a range of downstream tasks. This paper uses *Kinetics-400 and Kinetics-600* datasets for evaluation. Further, other available Audio-Visual datasets for emotion detection are *Aff-Wild2*, *MELD (Multimodal Emotion Lines Dataset)*, and *SEWA*.

*Table 2.1  
Emotion from different Domains / Modalities*

Modality	Features	Model	Dataset	Authors
Text	Millions of tweets (text) with emojis	'DeepMoji' model: Deep learning (SOTA architecture)	Tweets IEMOCAP MELD, DailyDialog EmoryNLP	Felbo, B., 2017 MIT
	Conversations	Directed Acyclic Graph based Network (DAG-ERC)   Transformer based LM		Shen, W. et al., 2021
	Conversations	Dialoguernn: Attentive RNN		Majumder et al., 2019
Audio	Mel Frequency Cepstral Coefficients (MFCC), Log-mel spectrograms, etc.	Bi-directional PCA, Linear Discriminant Analysis (LDA), Radial Basis Func (RBF)	eINTERFACE'05 RML	Ooi et al., 2014
Speech	Five-fold cross validation	Deep CNN	RAVDESS	Issa et al, 2020
	Non-linear features	SVM, KNN, LDA, Naive Bayes, Gradient Boosting, Random Forest		Krishnan et al., 2021
Sensors	On-Body sensor data (EEG, GSR); Accel, Environmental (Location, Audio, Weather)	CNN and LSTM	Custom	Kanjo et al., 2016
Vision	Facial expression	Attentional Convolutional Network	FER2013, CK+, JAFFE, FER6	Akhand et al., 2021 Minaee et al., 2021
Video	Ordered sequence of frames	CNNs, convolutional temporal feature LSTM, I3D (Inflated 3D convolution) Separable spatial, temporal (S3D) conv.	KINETICS	Yue-Hei Ng et al., 2015 Carreira et al., 2017 Xie et al., 2018
Multimodal Representations - Video and audio - Audio-Visual (AV)	Video, Audio, Text Audio-Visual	Video-Audio-Text Transformer (VATT), BERT, without CNN layers	Kinetics-400 Kinetics-600 (AV: Aff-Wild2, MELD, SEWA)	Akbari et al., 2021

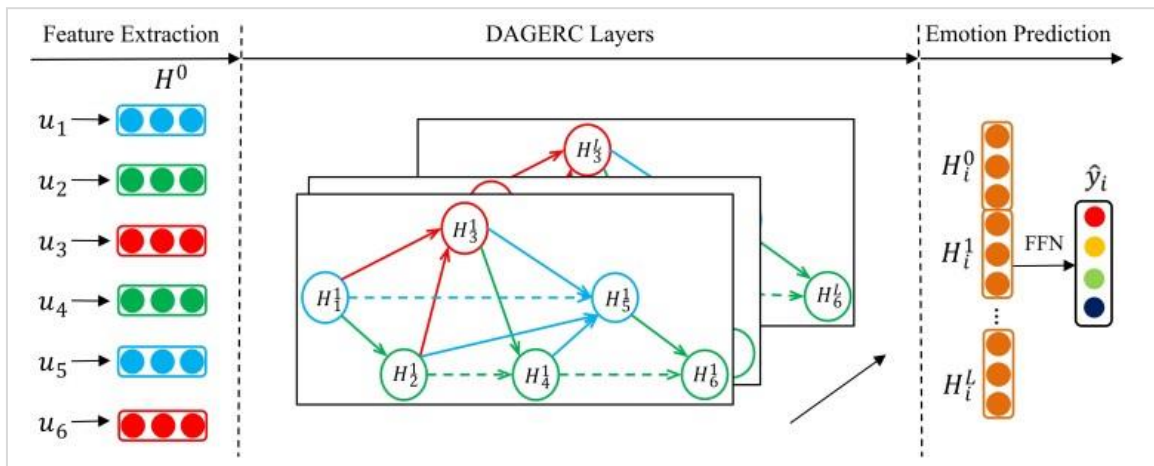
To summarize, in Human-Computer Interaction (HCI), human emotion recognition is vital in improving the overall user experience. As stated by Brave et al. (2007), emotion is a fundamental component of being human and motivates action and adds meaning and richness to virtually all human experiences. Emotion recognition has been pursued utilizing inputs from a variety of domains. Some research focuses on audio -based emotion identification (Ooi et al., 2014). Audio emotion recognition systems extract features like MFCC, Log-mel spectrograms, etc. from signals to determine the emotions. Another widely researched area is visual features-based emotion recognition. Vision-based systems like Akhand et al., (2021) use facial expression features to detect the subject's emotion in the image. Studies like You et al., (2016) focus on identifying the emotions evoked due to a particular image for the spectator. Other popular modalities from which human emotion is recognized is sensor-based techniques (Kanjo et al., 2019) and text (Majumder et al., 2019). The abundant availability of data also follows the wide research in these fields.

Some of the popularly used datasets include MELD (Revelle et al., 2009), IEMOCAP (Busso et al., 2008), RAVDESS (Livingstone et al., 2018) and many more.

## 2.3 Literature Review Details: Emotion Context from various Modalities

### 2.3.1 Textual Emotion Recognition – Directed Acyclic Graph

Figure 2.1  
Conversational Emotion Recognition with Directed Acyclic Graph Neural Network



Source: Shen et al. (2021)

Directed Acyclic Graph Network for Conversational Emotion Recognition (DAG-ERC): In this work (Shen et al., 2021), the utterances are encoded with a Directed Acyclic Graph (DAG) and they design a directed acyclic neural network, as shown in Figure 2.1, namely DAG-ERC to better model the intrinsic structure within a conversation. This work is inspired by DAGNN (Thost et al., 2021). DAG-ERC regards each utterance as a graph node, the feature of which can be extracted by a pre-trained Transformer-based Language Model. The pre-trained language model is firstly fine-tuned on each dataset, and its parameters are then frozen while training DAG-ERC. Then, they employ RoBERTa-Large

(Liu et al., 2019) as a feature extractor. The F1-Score for Interactive Emotional Dyadic Motion Capture (IEMOCAP), DailyDialog, Multimodal Emotion Lines Dataset (MELD), EmoryNLP Datasets are captured in the table below.

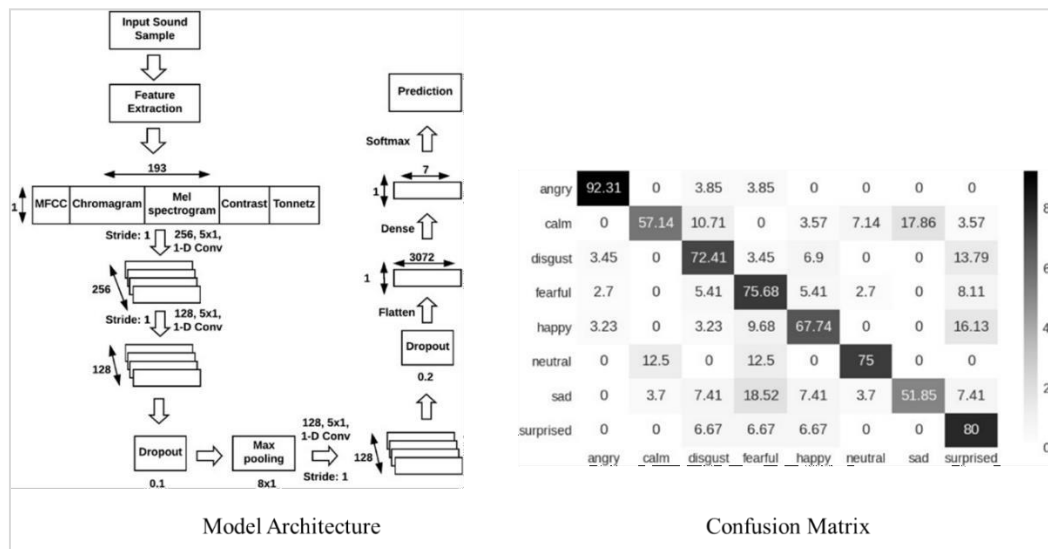
*Table 2.2  
Results Directed Acyclic Graph Network for Conversational Emotion Recognition*

Model	IEMOCAP	DailyDialog	MELD	EmoryNLP
DAG-ERC	68.03	63.65	59.33	39.02

### 2.3.2 Audio Emotion Recognition – Convolutional Neural Network

Issa et al. (2020) proposes SOTA model for emotion detection on Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) Dataset. The model used in this work is a Convolutional Neural Network (CNN) as shown in the Figure 2.2.

*Figure 2.2  
Speech Emotion Recognition with Deep CNN*



Source: Issa et al. (2020)

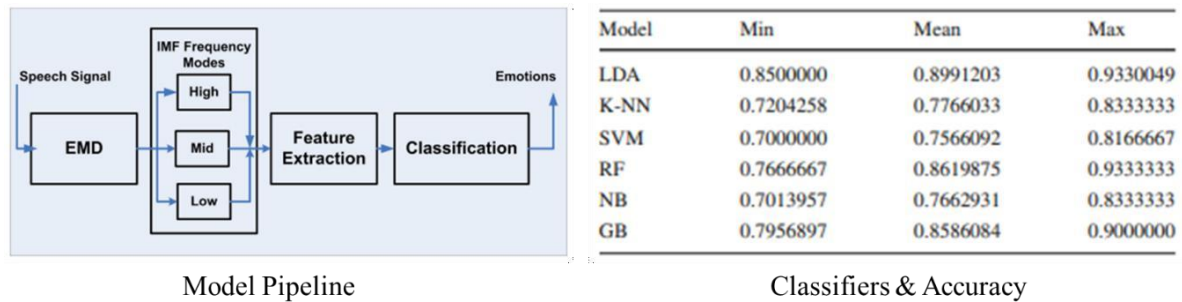
Five-fold cross validation is used to train the model in which the dataset was randomly divided into five groups of equal size and 80% of the dataset was used as training while the rest 20% was used for testing. This schema is performed five times and the average classification accuracy on test sets is computed. Since data partitioning is performed randomly, the classification is speaker-independent. The model is trained for a total of 700 epochs and obtains 71.61% for RAVDESS with 8 classes, as shown in the confusion matrix in the Figure 2.2.

### 2.3.3 Audio Emotion Recognition – Empirical Mode Decomposition

The next work uses Empirical Mode Decomposition (EMD) for audio emotion recognition, and the model pipeline of this work is shown in Figure 2.3.

Figure 2.3

Audio Emotion Detection with Empirical Mode Decomposition on TESS Dataset



Source: Krishnan et al. (2021)

Emotion classification from speech signal based on empirical mode decomposition and non-linear features: Speech emotion recognition – This work is proposed by Krishnan et al. (2021). Empirical Mode Decomposition (EMD) and non-linear features are used as the input to the proposed model. The emotions are recognized from speech signals by decomposing them into intrinsic mode functions (IMF). Later, five unique randomness

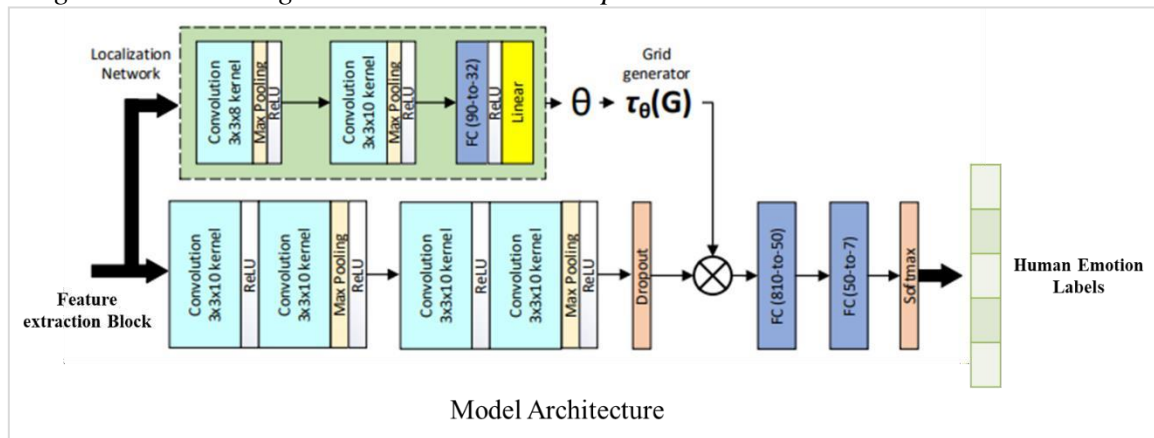
measures are computed through entropy measures and SOTA ML classifiers are trained on the entropy features. The classifiers compared are SVM, LDA, Naïve Bayes, KNN, Gradient Boosting, and Random Forest using the Mean Balanced Accuracy (MBA) metric, and captured in Figure 2.3. Out of all the classifiers, LDA performs the best and brings the accuracy of 0.93 on Toronto emotional speech set (TESS) dataset.

### 2.3.4 Image Emotion Recognition using Attentional Convolutional Network

Minaee et al. (2019) proposes this work: Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. The model architecture of this work is captured in Figure 2.4. The following shows the summary of this work:

- Built on observation that not all the parts of the face are important for detecting specific emotion. Thus, it utilizes the feature extraction part and spatial transformer (localization network) to focus on important facial region.
- After regressing the transformation parameters  $\tau$ , the input was transformed to sampling grid  $T(\theta)$ , producing the warped data

Figure 2.4  
Image Emotion Recognition - Architecture Pipeline



Source: Minaee et al. (2019)

- Network utilizes few numbers of layers for faster inference speed and is more suitable for real time applications.
- The Emotion classification accuracy on facial expression recognition (FER), Japanese Female Facial Expression (JAFFE), and the extended Cohn-Kanade dataset (CK+) dataset is captured in the table.

*Table 2.3  
Emotion Classification Accuracy*

<b>Dataset</b>	<b>Training Samples Used</b>	<b>No. of Test Samples</b>	<b>Emotion Classification Accuracy</b>
FER-2013	34,000	7000	99.3%
JAFFE	120	70	92.8%
CK+	420	113	98%

Further, the Architecture Pipeline Modification that can be planned are as follows:

- Experimenting with different transformations to warp the input to output images
- Using multiple feature extraction blocks and combining the output for getting more refined representation
- Utilizing the representation from Visual Geometry Group (VGG), ResNet18 for this downstream task. (ResNet is a Residual Network, and ResNet-18 is a CNN network that is 18 layers deep).

Image Data set for Emotion Recognition:

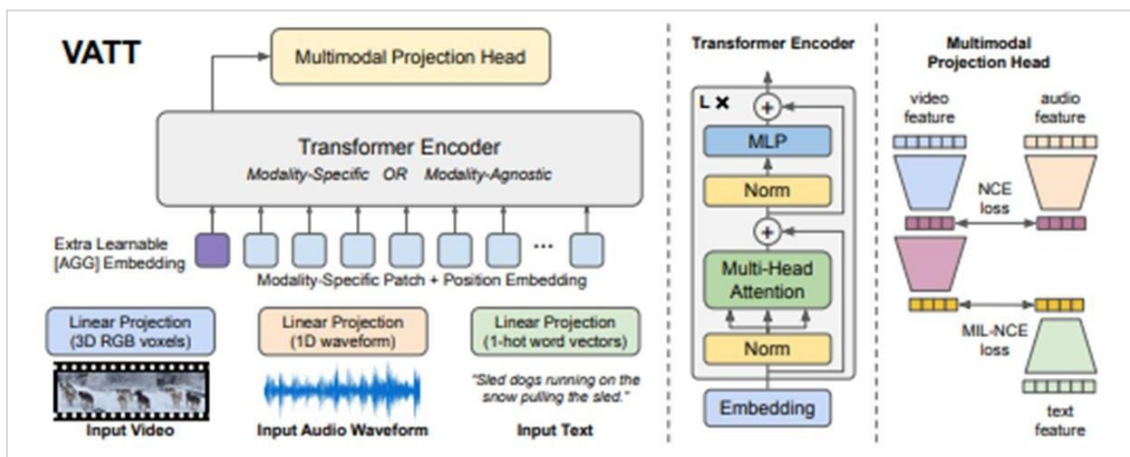
- FER2013: 35,887 images of 48x48 resolution
- CK+: 539 sequences across 123 subjects
- JAFFE: 213 images of 7 facial expressions

- FERF (Facial Expression Research Group Database): 55,767 annotated face images of 6 stylized characters

### 2.3.5 Video Emotion Recognition – Video-Audio-Text Transformer

- Video-Audio-Text Transformer (VATT)
- Akbari et al (2021) proposes VATT: Transformers for multimodal self-supervised learning from raw Video Audio and Text. The model architecture details are captured in Figure 2.5 (Akbari et al, 2021).
- The backbone network comprises of regular Transformer Architecture (Bidirectional Encoder Representations from Transformers - BERT). No CNN Layers are used. This makes it more suitable for multimodal data. Text input can also be added if required.
- Since unsupervised learning (contrastive losses) is utilized, we can benefit from large dataset of unlabeled videos available (Ex: YouTube videos)

Figure 2.5  
Video / Multimodal Transformer Architecture for Human Emotion Classification



Source: Akbari et al (2021)



- Such a pretrained network which has learnt the (semantically aware) common representative from video and audio data can be used for the range of downstream task.
- VATT Transformer achieves top-1 accuracy of 82.1% on Kinetics-400, 83.6% on Kinetics-600 and 41.1% on Moments in Time for Video action detection. This is better than current state-of-the-art (SOTA) architectures.
- This also sets a new record on audio event recognition by achieving the mAP of 39.4% on Audio dataset.

In addition, the following modifications can be done to experiment human emotion classification on multimodal data:

- Audio along with the video would be more suitable for human emotion classification.
- After training the transformer on larger corpus of video, we can add a classification layer for human emotion classification. This layer takes an input, the vectors from common video-audio subspace.
- In this new model, the pretrained backbone network can either be frozen or trained from previous weights based on the size of dataset.

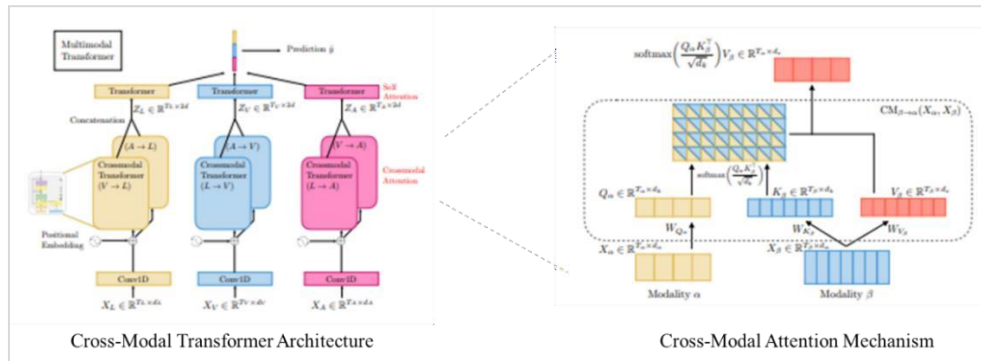
Audio-Visual Dataset for Emotion Detection:

- AFF-Wild2 (Affect Detection in-the-wild database): 558 videos, 2.8 Million frames
- MELD (Multimodal Emotion Lines Dataset): Multimodal Emotional Line Dataset (1400 dialogues)
- SEWA (Automatic Sentiment Analysis in-the-Wild database): 1525 minutes of audio-visual data of people's reaction



Figure 2.7

Multimodal Emotion Recognition: Cross-Modal Transformer Architecture & Attention



Source: Tsai et al (2019)

### 2.3.7 Video Emotion Recognition – An identity-free video dataset

- Liu et al (2021) proposes IMiGUE: An Identity-free video dataset for Micro-gesture Understanding and Emotion Analysis
- This paper makes 2 main contributions towards Emotion understanding through micro-gestures.
- Current gesture datasets have 2 main problems. The gesture for emotion is acted out an don't natural, and hence might not be relevant in the real world. Also, identification through bio-metric data and thus **user privacy** is a concern.

Figure 2.8

Video Dataset – identity free micro-gestures (Cover-face, fold-arms, cross-fingers)

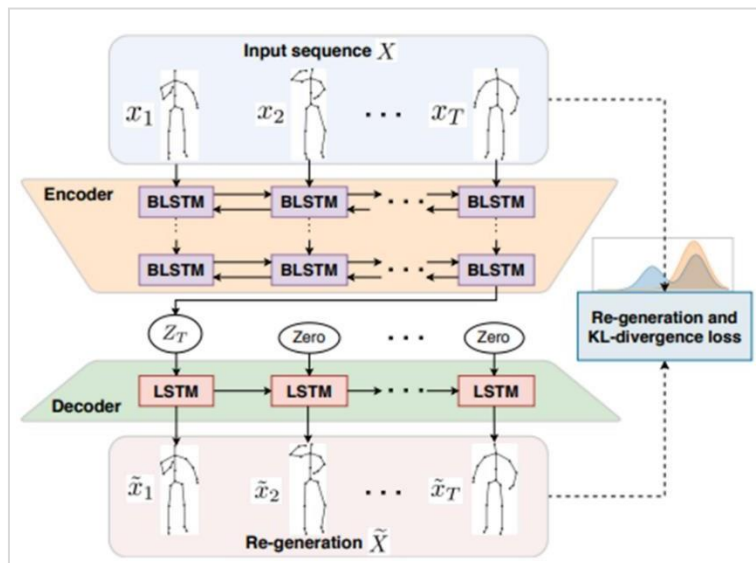


Only Pose (2D skeleton) is captured

Source: Liu et al (2021)

- The Figure 2.8 (Liu et al, 2021) shows how User privacy is maintained, when emotions are analyzed using micro-gestures, which might prove to be a desired property in case of sensitive emotions / conditions like depression. Further, since only a pose (2D skeleton) is required, the user privacy is maintained, as shown in the figure.
- Based on psychological research, the authors have defined 32 micro-gestures, which includes touching ears, folding arms, shrugging shoulders. Tennis Grand Slam post-match conference is then annotated at clip level (micro-gestures) and at the video level (positive or negative) emotion, based on whether player won or lost the match.

Figure 2.9  
Video Emotion Recognition – Unsupervised Encoder-Decoder Network



Source: Liu et al (2021)

- The Figure 2.9 (Liu et al, 2021) summarizes the architectural contribution. As shown in the figure, the unsupervised encoder-decoder network is proposed. After training, encoder state for the input sequence is used to classify using K-Nearest Neighbor (KNN) algorithm.

### 2.3.8 Sensor based Emotion Recognition

Park et al (2020) proposes K-EmoCon Dataset: A multimodal Sensor Dataset for continuous Emotion Recognition in Naturalistic Conversations.

Description:

- A multimodal dataset with comprehensive annotations of continuous emotions during naturalistic conversation (Dyadic Communication).
- Participants annotated emotion with 20 unique categories.

#### **Modalities and Bio-signals measured:**

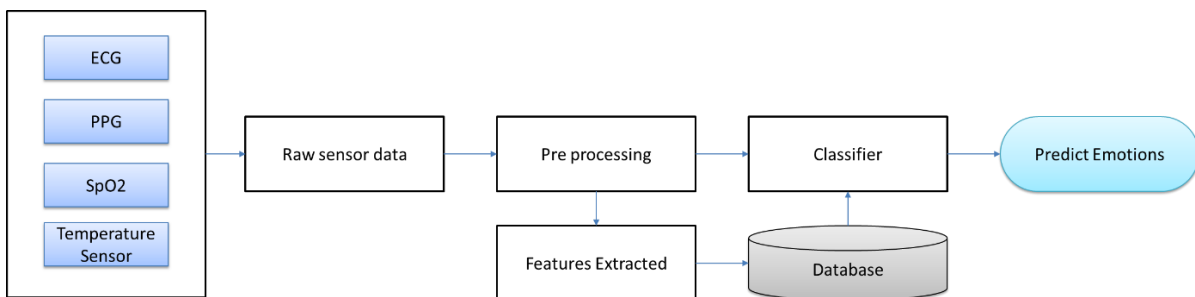
- Videos (Face, Gesture) and Audio: By two Samsung Galaxy S7 smartphone cameras
- Blood Volume Pulse (BVP): Measured by a photo-plethysmo-graphy (PPG) sensor and detect heart rate using electrocardiogram (ECG) sensor.
- Electroencephalogram (EEG) signals: Acquire brain signals via 2 dry sensor electrodes.
- Body Temperature, Electrodermal activity (EDA): Measures conductance of the skin in response to sweat segregation
- The paper does not show any architecture, hence, a brief overview of how emotions can be predicted is captured in the figure 2.10 below.

#### **Emotion Annotation Categories:**

- Arousal, Valence: 1 (Very low) – 5 (Very high)
- Cheerful, Happy, Angry, Nervous, Sad: 1 (very low) – 4 (very high)
- Common BROMP Affective categories: Boredom, Confusion, Delight, Engaged Concentration, Frustration, Surprise

- Less Common BROMP affective categories: Confrustion, Contempt, Dejection, Disgust, Eureka, Pride, Sorrow
- No. of Emotion Annotations: Number of 5 second intervals annotated: 4,159 (self), 4,159 (partner), 20,803 (five-external observers)

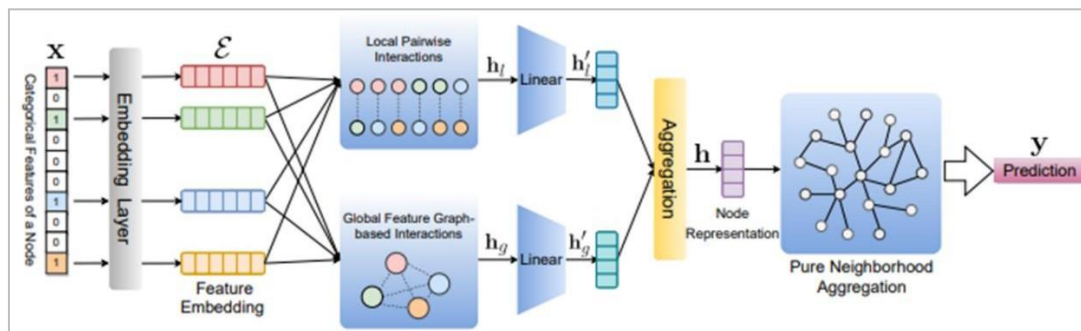
Figure 2.10  
Sensors based Emotion Prediction



### 2.3.9 Emotion Prediction from Categorical Data

- Most existing work linearly combines the embeddings of the node features
- Feature interaction is important especially in categorical features

Figure 2.11  
Categorical Emotion Prediction - Graph Convolutional Network



Source: Chen et al. (2021)

- Chen et al. (2021) proposes CatGCN: Graph Convolutional Networks with Categorical Node features. The model architecture is captured in Figure 2.11.
- This work CatGCN is proposed for Graph learning on categorical features.
- This uses Local interaction modeling on each pair of node features and global interaction modeling
- It further refines the enhanced initial node representations with the neighborhood aggregation-based graph convolution
- The work conducted extensive experiments on three tasks of user profiling (the prediction of user age, city and purchase level) from Tencent and Alibaba datasets.

### 2.3.10 Rule based ML: Representation Learning

Wang et al. (2021) proposes Scalable rule-based representation learning for interpretable classification. This work is proposed to enable Rule based ML. The sample architecture from this work is shown in Figure 2.12. The explanation for this work is explained below:

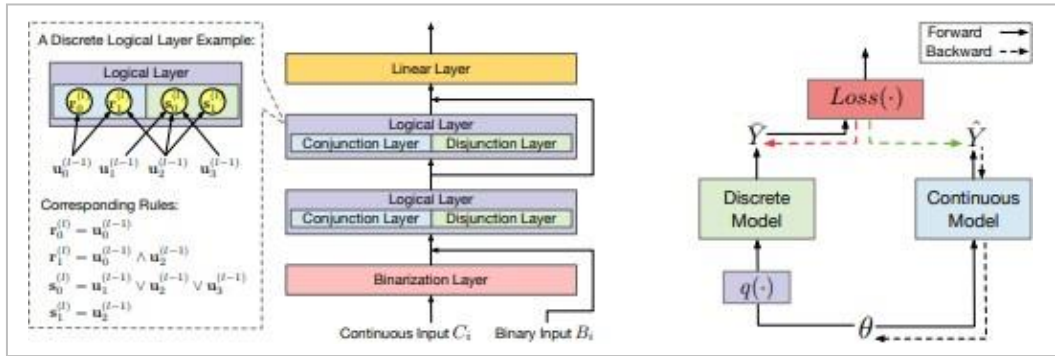
- Class decisions depending on various ‘if-then’ conditions
- Rule based classifiers are generally not mutually exclusive
- Decision boundaries created by them are linear

How can the overlapping rule problem be handled?

- Either rules can be **ordered** i.e., the class correspondence to the highest priority rule triggered is taken as the final class. Or, we can assign **votes** for each class depending on some of their weights i.e., the rules remain unordered.
- Example - **Rules:** Odour = pungent and habitat = urban → class = poisonous
- The example is illustrated in detail as shown below:

R1: (Give birth = no)  $\wedge$  (Can fly = yes)  $\rightarrow$  Birds  
 R2: (Give birth = no)  $\wedge$  (Live in water = yes)  $\rightarrow$  Fishes  
 R3: (Give birth = yes)  $\wedge$  (Blood type = warm)  $\rightarrow$  Mammals  
 R4: (Give birth = no)  $\wedge$  (Can fly = no)  $\rightarrow$  Reptiles  
 R5: (Live in water = sometimes)  $\rightarrow$  Amphibians

Figure 2.12  
 Rule-based Representation Learning – Example



Source: Wang et al. (2021)

**2.3.11 Image Emotion Recognition – Context based Emotion Recognition**

Kosti et al (2019) proposes Context based Emotion Recognition using emotic dataset. Emotic dataset is named after EMOTions In Context.

**Model Description:**

The basic objective of this model structure is to bring in context along with the body posture / facial expression while classifying an image for emotion. The model gives 2 types of emotion outputs:

- One is getting a discrete categorical classification



- Second is predicting the values on continuous Valence-Arousal-Dominance (VAD) model

Twenty-Six discrete categories (Kosti et al, 2019) have been used keeping in mind the concept of visual separability of emotions, and 3 dimensions (VAD) have been used for continuous models.

- Valence (V): How positive or pleasant the person is.
- Arousal (A): Measures the agitation level of the person (calm to agitated)
- Dominance (D): Measures the level of control a person shows (submissive to dominant)

*Table 2.4  
Twenty-six Discrete Emotion Categories*

1. <b>Affection:</b> fond feelings; love; tenderness
2. <b>Anger:</b> intense displeasure or rage; furious; resentful
3. <b>Annoyance:</b> bothered by something or someone; irritated; impatient; frustrated
4. <b>Anticipation:</b> state of looking forward; hoping on or getting prepared for possible future events
5. <b>Aversion:</b> feeling disgust, dislike, repulsion; feeling hate
6. <b>Confidence:</b> feeling of being certain; conviction that an outcome will be favorable; encouraged; proud
7. <b>Disapproval:</b> feeling that something is wrong or reprehensible; contempt; hostile
8. <b>Disconnection:</b> feeling not interested in the main event of the surrounding; indifferent; bored; distracted
9. <b>Disquietment:</b> nervous; worried; upset; anxious; tense; pressured; alarmed
10. <b>Doubt/Confusion:</b> difficulty to understand or decide; thinking about different options
11. <b>Embarrassment:</b> feeling ashamed or guilty
12. <b>Engagement:</b> paying attention to something; absorbed into something; curious; interested
13. <b>Esteem:</b> feelings of favourable opinion or judgement; respect; admiration; gratefulness
14. <b>Excitement:</b> feeling enthusiasm; stimulated; energetic
15. <b>Fatigue:</b> weariness; tiredness; sleepy
16. <b>Fear:</b> feeling suspicious or afraid of danger, threat, evil or pain; horror
17. <b>Happiness:</b> feeling delighted; feeling enjoyment or amusement
18. <b>Pain:</b> physical suffering
19. <b>Peace:</b> well being and relaxed; no worry; having positive thoughts or sensations; satisfied
20. <b>Pleasure:</b> feeling of delight in the senses
21. <b>Sadness:</b> feeling unhappy, sorrow, disappointed, or discouraged
22. <b>Sensitivity:</b> feeling of being physically or emotionally wounded; feeling delicate or vulnerable
23. <b>Suffering:</b> psychological or emotional pain; distressed; anguished
24. <b>Surprise:</b> sudden discovery of something unexpected
25. <b>Sympathy:</b> state of sharing others emotions, goals or troubles; supportive; compassionate
26. <b>Yearning:</b> strong desire to have something; jealous; envious; lust

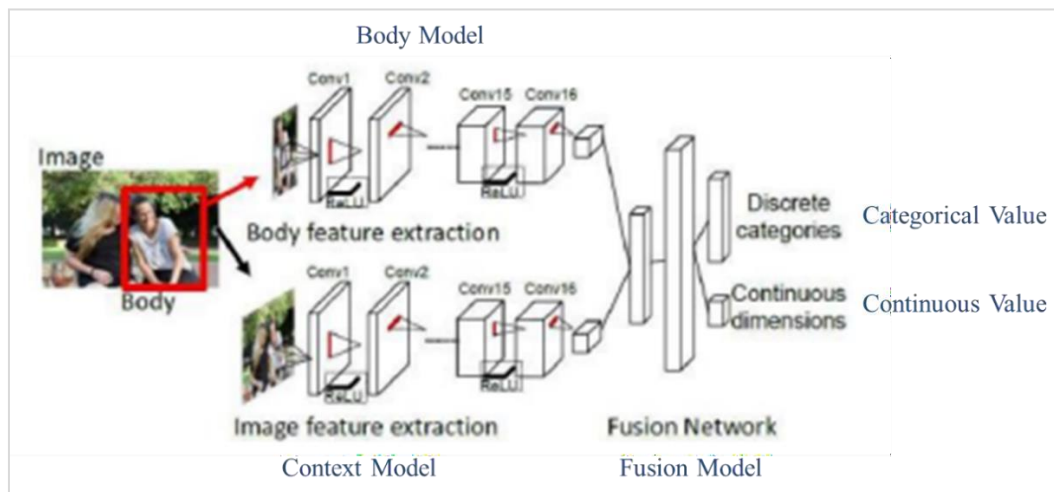
Source: Kosti et al (2019)

The model structure (Kosti et al, 2019) consists of 3 models:

1. Body Model:

- This model is pretrained on image net dataset. The output of this model are various body cues that can hint towards the emotion shown by the person in the image.
- This model is based on ResNet18 architecture, where the last layer of ResNet18 has been dropped.
- The input of this model is the portion of the image centered towards the body of the person, a (3, 224, 224) input is given and an output of (512, 1) is obtained.

Figure 2.13  
Context based Emotion Recognition



Source: Kosti et al (2019) and Adapted.

2. Context Model

- This model is pretrained on places dataset. This model gives output regarding the various context features that can be associated with various emotions.
- It is based on ResNet18 architecture where the last layer of the architecture is dropped

- The input of this model is that the whole RGB image is resized to the shape of (3, 224, 224), and an output of (512, 1) is obtained.

### 3. Emotic Model (Fusion Model)

- This is a fusion model that takes the output of the body and context model as its input, and generates 2 output vector, one for categorical values and second for continuous values.
- The model is 3 layers deep with an input of (1024, 1) and an output of (26, 1) & (3, 1)

#### **Model status – Quick experiment:**

- Able to recreate and train the model (PyTorch) on Emotic Dataset
  - Average precision: 28 (Research paper claimed 29.45)
  - ROM: Body, Context : 44.8 MB each, Fusion: 1.1 MB
- Tflite conversion has drastic reduction in accuracy
- Converted to PyTorch Lite and able to deploy on device
  - Average precision: 20.4
  - ROM: Body, Context: 89.4 MB each, Fusion: 2.2 MB
  - Init time for first launch: ~20 sec, and for sub-sequent launches: ~6 sec.
  - Inference time: ~1 sec.

#### **2.3.12 Self Supervised Learning (SSL)**

For any supervised task, availability of good annotated data is a primary requirement. However, collection and annotation of data is a very costly and time-consuming process. Hence, we need a method to get this annotated data quickly and cheaply, SSL provides us with a way to get this done.

SSL method can be divided into 2 sections:

1. Designing and training on pretext task
2. Training on downstream task

The following method is being explored to be used as a pretext task.

### **Representation Learning by solving Jigsaw Puzzles:**

Representation Learning is a method which involves utilization of meta data, or the images itself to learn feature representation of object present in the image. This reduces the time and cost of labelling the data by a huge margin. One such method is by designing a model to solve Jigsaw puzzle. The method is discussed below.

Firstly, an image that needs to be passed to the model is cut into 9 pieces and each piece is labelled from 1~9. Then, the pieces are shuffled with different set of permutations. The task of the model is to predict which piece number corresponds to which position. While doing so, the model not only learns the feature representation but the semantic representation as well for any given image.

The image is passed to the model a number of times for different permutations and the final results are quite satisfactory. This method can be implemented to get good features of the subject without explicitly labelling each image.

A context Free network is used in the model. As compared with AlexNet (Top 1 accuracy of 57.4%), this model gives the Top-1 accuracy of 57.1% with much less parameters to learn and huge reduction in training time.

## **2.4 Literature Review Summary: Data collection**

Data collection is the first and one of the most crucial steps in performing any analysis. Especially in the field of Machine learning (ML) and Artificial Intelligence (AI), the quality and depth of data will determine the level of AI applications that can be achieved. Data collection and analysis methods have been explored since the early 1600s when John Graunt (1977) conducted data analysis on gender-based death rates and attempted to predict life expectancy. Over the years, the research in this field evolved where questionnaire-based data collection was explored (Baker, 2003). Recently, due to the advent of smartphone usage, UI designs for more effective data collections have been extensively studied (Schobel, 2014).

With the advent of smartphone usage, one of the most significant indicators of a user's emotional state is the user-activity data. Although extensive research in emotion recognition utilizes various modalities as listed above, research in this field of predicting the mood/emotion of users using mobile-activity data, which is mostly categorical, is sparse. One important blocking factor is the lack of data. Previous work in this area includes MoodScope (LiKamWa et al., 2013). Here they create a data collection application and collect information like SMS, email, phone call, application usage, web browsing, and location from 32 participants four times a day to build statistical usage models to estimate mood. They use least-squares multiple linear regression to perform the modelling, along with Sequential Forward Selection (SFS). Bogomolov et al. (2014) and Hung et al. (2016) focus on detecting negative emotions like stress, depression from user activity data collected using a custom application. Since labelled data of this kind is sparse, iSelf (Sun et al., 2017) attempts to tackle the problem of cold- start conditions to predict

emotion labels. Shapsough et al. (2016) focuses on emotion recognition using smartphones using sensor and keyboard usage data.

More interesting approaches to collect labelled data using smartphones have been explored in the following years. MoodExplorer (Zhang et al., 2018) proposes the system framework for compound emotion detection via smartphone sensing. They collect data from 30 university students and apply feature selection techniques to detect compound emotions rather than singular labels. Morshed et al. (2019) propose a framework for predicting mood instability using passively sensed data gathered from smartphones and wearables. Authors formulated mood prediction as regression problem and considered several modalities based on audio, sensor, GPS information etc. to predict mood instability. Darvariu et al. (2020) developed MyMood, a smartphone application that allows users to periodically log their emotional state together with pictures from their everyday lives along with passive sensor data. They use this visual information along with sensor data to develop deep learning methods for human emotion prediction. Roshanaei et al. (2017) developed the Emosensing app to collect ground truth data for emotion prediction, which requires the user to launch the application and log the emotion manually. They incentivize users with monetary benefits to encourage data logging. Further, they used the collected data: activity, smartphone app usage, and location (private data) to predict 13 different user emotions. Further studies also use personality information and relate to users' emotional states, as studied in (Donovan et al., 2021).

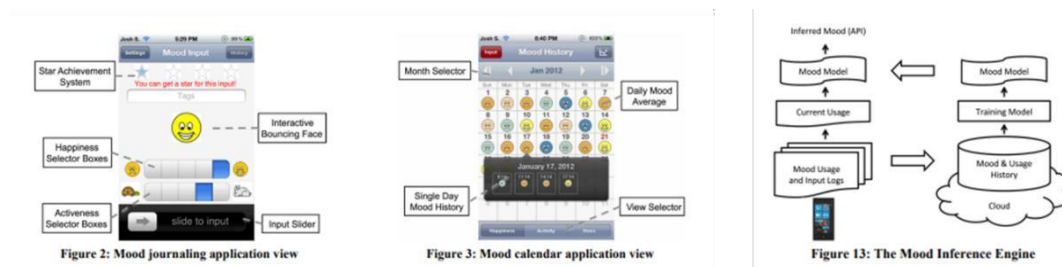
Table 2.5  
Methods for Data Collection

Paper/App	Data collection techniques	Participants	Authors
MoodScope	Created data collection app, and collect private information like SMS, email, phone call, application usage, web browsing, and location	32 participants four times a day	LiKamWa et al., 2013
MoodExplorer	Collect labelled data using smartphones (Private content)	30 university students	Zhang et al., 2018
EmoSensing	Collect ground truth data for emotion prediction, which requires the user to launch the application and log the emotion manually. They incentivize users with monetary benefits to encourage data logging. Used activity, smartphone app usage, and location (private data) to predict 13 different user emotions	27 students	Roshanaei et al., 2017
Prediction of Mood Instability with Passive Sensing	<a href="http://studentlife.cs.dartmouth.edu/dataset.html">http://studentlife.cs.dartmouth.edu/dataset.html</a> : sensing  -activity  -audio  -conversation  -bluetooth  -dark  -gps  -phonecharge  -phonelock  -wifi  -wifi_location	NA	Morshed et al., 2019
MyMood	Allows users to periodically log their emotional state together with pictures from their everyday lives along with passive sensor data	Everyday	Application

## 2.5 Literature Survey Details: Emotion from Categorical Data

### 2.5.1 MoodScope: Building a Mood Sensor from Smartphone Usage Patterns

Figure 2.14  
MoodScope: Use case and Engine



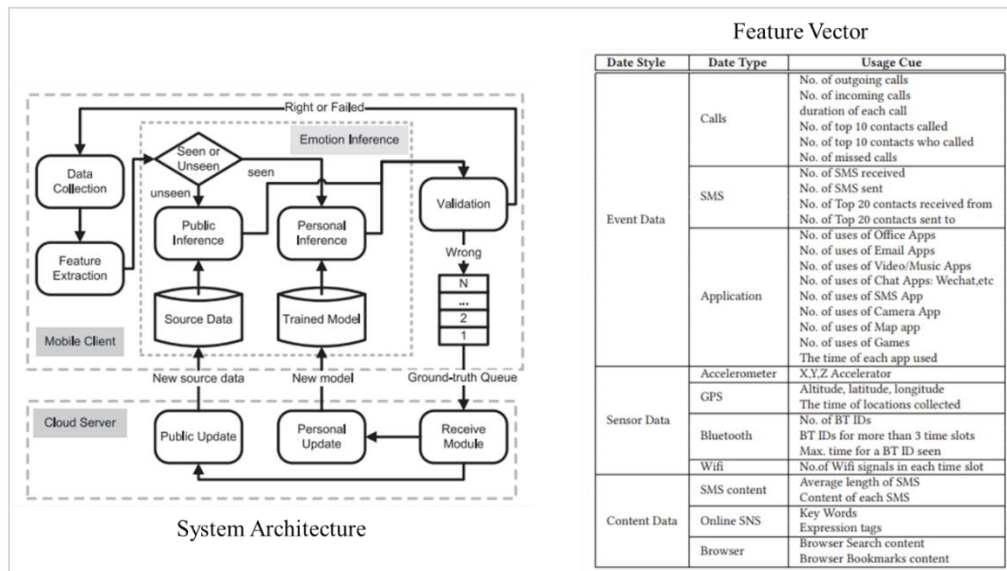
Source: LiKamWa et al. (2013)

- Data Collected from 32 participants, 4 times a day, with at-least 3 hours between each input
- Using only 6 piece of usage information namely SMS, Email, Phone call, Application usage, Web browsing, Location, and Build statistical usage models to estimate Mood
- Use Least-square multiple Linear Regression to perform modeling and Sequential-Forward Selection (SFS)

- Use cases targeted / mentioned: Recommendation (Netflix, spotify); Social networks by allowing users to share mood states automatically; to log user moods and browse mood history, as shown in Figure 2.14 (LiKamWa et al., 2013)

## 2.5.2 iSelf: Towards Cold-start Emotion Labeling using Transfer Learning with Smartphones

Figure 2.15  
iSelf: System Architecture and Feature Vector



Source: Sun et al. (2017)

- Systems which can infer personal emotions automatically in cold-start conditions, i.e., with only a few labelled samples on smartphones.
- Three kinds of data: Event data (ex: calls and applications), Sensor data (ex: wi-fi), and Content data (SMS content), as shown in Figure 2.15 (Sun et al., 2017)
- 3600 samples of 10 participants during 30 days, collected for the duration of 1 hour



- Use Support Vector Machine (SVM) for training personal model

### 2.5.3 Daily Stress Recognition from Mobile Phone Data, Weather Conditions and Individual Traits

Figure 2.16  
Daily Stress Recognition: Basic Features and Bluetooth Proximity Features

List of Basic Features	List of Basic Bluetooth Proximity Features
<p><u>General Phone Usage</u></p> <ol style="list-style-type: none"> <li>1. Total Number of Calls (Outgoing+Incoming)</li> <li>2. Total Number of Incoming Calls</li> <li>3. Total Number of Outgoing Calls</li> <li>4. Total Number of Missed Calls</li> <li>5. Number of SMS received</li> <li>6. Number of SMS sent</li> </ol> <p><u>Diversity</u></p> <ol style="list-style-type: none"> <li>7. Number of Unique Contacts Called</li> <li>8. Number of Unique Contacts who Called</li> <li>9. Number of Unique Contacts Communicated with (Incoming+Outgoing)</li> <li>10. Number of Unique Contacts Associated with Missed Calls</li> <li>11. Entropy of Call Contacts</li> <li>12. Call Contacts to Interactions Ratio</li> <li>13. Number of Unique Contacts SMS received from</li> <li>14. Number of Unique Contacts SMS sent to</li> <li>15. Entropy of SMS Contacts</li> <li>16. Sms Contacts to Interactions Ratio</li> </ol> <p><u>Active Behaviors</u></p> <ol style="list-style-type: none"> <li>17. Percent Call During the Night</li> <li>18. Percent Call Initiated</li> <li>19. Sms response rate</li> <li>20. Sms response latency</li> <li>21. Percent SMS Initiated</li> </ol> <p><u>Regularity</u></p> <ol style="list-style-type: none"> <li>22. Average Inter-event Time for Calls (time elapsed between two events)</li> <li>23. Average Inter-event Time for SMS (time elapsed between two events)</li> <li>24. Variance Inter-event Time for Calls (time elapsed between two events)</li> <li>25. Variance Inter-event Time for SMS (time elapsed between two events)</li> </ol>	<p><u>General Bluetooth Proximity</u></p> <ol style="list-style-type: none"> <li>1. Number of Bluetooth IDs</li> <li>2. Times most common Bluetooth ID is seen</li> <li>3. Bluetooth IDs accounting for n% of IDs seen</li> <li>4. Bluetooth IDs seen for more than k time slots</li> <li>5. Time interval for which a Bluetooth ID is seen</li> <li>6. Entropy of Bluetooth contacts</li> </ol> <p><u>Diversity</u></p> <ol style="list-style-type: none"> <li>7. Contacts to interactions ratio</li> </ol> <p><u>Regularity</u></p> <ol style="list-style-type: none"> <li>8. Average Bluetooth interactions inter-event time (time elapsed between two events)</li> <li>9. Variance of the Bluetooth interactions inter-event time (time elapsed between two events)</li> </ol>

Source: Bogomolov et al. (2014)

- 2-class classification problem (Non-stressed vs Stressed) based on information concerning different type of data:
  - People activities as detected through smart phones
  - Weather conditions
  - Personal traits
- Uses an ensemble of tree classifiers based on a Random Forest algorithm. Data set capturing the lives of 117 subjects for 6 months.
- Collected weather and proximity data as well

- List of features used for collection/training are captured in Figure 2.16 (Bogomolov et al, 2014). Further, the work uses Principal Component Analysis (PCA) and Pearson correlation for feature selection.

### 2.5.4 Does Smartphone Use drive our emotions or vice versa? A Casual Analysis

Figure 2.17

Smartphone Application Launch & Usage and Emotions

Metric	Contempt	Disgust	Joy	Sadness	Surprise
Total Apps Launch	0.203 (0.140)	0.174 (0.205)	0.194 (0.132)	0.108 (0.214)	0.202 (0.311)
Communication Apps Launch	0.226 (0.261)	0.059 (0.414)	0.147 (0.224)	-0.002 (0.391)	0.260 (0.383)
Social Apps Launch	0.305 (0.403)	0.092 (0.429)	0.479 (0.153)	-0.083 (0.555)	-0.015 (0.288)
Work Apps Launch	0.484 (0.185)	0.251 (0.129)	0.257 (0.252)	0.070 (0.555)	0.088 (0.589)
Entertainment Apps Launch	0.202 (0.348)	0.239 (0.542)	0.487 (0.481)	0.262 (0.847)	-0.151 (0.621)

Table 1: Effect Sizes (and SD) for Causality between Application Launch and Emotions. Blue: phone use drives emotions; Orange: emotion drives phone use.

Metric	Contempt	Disgust	Joy	Sadness	Surprise
Total Apps Usage Duration	0.091 (0.172)	0.047 (0.220)	0.138 (0.191)	0.086 (0.214)	0.144 (0.247)
Communication Apps Usage Duration	-0.118 (0.534)	0.131 (0.393)	0.124 (0.331)	0.122 (0.388)	0.177 (0.363)
Social Apps Usage Duration	0.221 (0.387)	0.093 (0.680)	0.129 (0.360)	0.168 (0.215)	-0.025 (0.469)
Work Apps Usage Duration	-0.060 (0.558)	0.064 (0.574)	-0.180 (0.316)	0.034 (0.547)	0.265 (0.315)
Entertainment Apps Usage Duration	0.149 (0.302)	-0.010 (0.616)	0.256 (0.471)	-0.050 (0.680)	-0.359 (0.379)

Table 2: Effect Sizes (and SD) for Causality between Application Usage Duration and Emotions. Blue: phone use drives emotions; Orange: emotion drives phone use.

Sarsenbayeva et al. (2020)

- Sarsenbayeva et al. (2020) explains the existence of a bidirectional causal relationship between smartphone application use and user emotions.
- In a 2-week long study with 30 participants captured 502,851 instances of smartphones application use with corresponding emotional data from facial expressions
- Bi-directional influence: Our experience causes emotional reactions and in turn these emotions shape our behavior and interactions
- Convergent Cross-Mapping (CCM) algorithm is used
- Show that, overall, phone use drives certain emotions rather than the other way around. Furthermore, identify specific application on categories which actually drives users' emotion.

## 2.5.5 Predicting Personality from patterns of behavior collected with smartphones

Figure 2.18  
Big 5 Personality Prediction Model

Table 1. Top five predictors per prediction model

Personality dimension	Top five predictors
O, openness	Daily mean length text messages   robust mean dur sports news apps   daily robust variation dur phone ringing   daily robust mean no. photos   robust mean dur sports news apps night
O2, openness to aesthetics	Robust mean dur sports news apps   daily mean no. photos   daily mean no. unique sports news apps   robust mean dur nightly sports news app   daily mean no. sports news apps
O3, openness to feelings	Excess music acousticness   daily mean no. unique sports news apps per week   robust variation dur shared transportation apps   daily robust variation in dur phone ringing   daily mean no. unique sports news apps
O4, openness to actions	Mean no. of phone ringing night   daily mean no. of ringing events   daily mean no. Google Maps   mean no. calls night   irregularity of phone ringing
O5, openness to ideas	Loudness fourth most listened song   robust mean dur sports news apps   daily SD no. of photos   robust mean dur <i>Süddeutsche Zeitung</i> (newspaper)   robust mean dur Samsung Notes
O6, openness to value and norm	Daily mean no. unique sports news week   daily mean no. Facebook   daily mean no. sports news   daily mean no. unique sports news weekend   daily mean no. Kicker (soccer news)
C, conscientiousness	Robust mean dur weather app night   daily SD sum interevent time   robust mean time last event   robust variation dur checkup monitoring apps   robust variation first event weekdays
C2, love of order	Daily SD sum interevent time   robust mean dur news-magazine apps   daily mean no. unique email apps   mean mean charge disconnection   robust variation dur TV-filmguide apps
C3, sense of duty	SD dur nightly downtime   robust mean time first event weekdays   robust variation time last event weekdays   robust mean dur Stadtwerke München Fahrinfo München (public transportation)
C4, ambition	Robust mean time first event   robust variation first event weekdays   robust mean time last event   robust variation time first event weekends   daily mean no. Google Playstore
C5, discipline	Robust variation time first event weekdays   robust mean time first event weekdays   robust mean dur weather apps night   robust variation time first event weekends   daily SD sum interevent time
C6, caution	Robust variation time last event weekdays   SD dur nightly downtime Sunday til Thursday   similarity contacts phone and messaging   robust variation time last event   mean music valence weekends
E, extraversion	Nightly mean no. phone ringing   nightly mean no. calls   daily mean no. outgoing calls   daily mean no. phone ringing   nightly mean no. outgoing calls
E1, friendliness	Daily mean no. phone ringing   irregularity of phone ringing weekend   daily SD no. incoming calls   daily robust variation sum dur phone ringing   daily SD sum dur incoming calls
E2, sociableness	Mean no. calls night   daily mean no. outgoing calls   mean no. phone ringing night   mean no. outgoing calls night   irregularity of phone ringing weekend
E3, assertiveness	Daily mean no. outgoing calls   daily mean no. contacts per week   daily mean no. contacts outgoing calls   daily mean no. contacts calls   mean no. calls night
E4, dynamism	Daily mean no. outgoing calls   mean no. phone ringing night   daily mean no. contacts outgoing calls   mean no. calls night   daily mean no. phone ringing
E5, adventurousness	Mean no. phone ringing night   mean no. calls night   irregularity of phone ringing   mean no. outgoing calls night   irregularity of calls
ES1, carefreeness	Daily mean no. Android-Email (app)   daily mean no. screen unlocks   robust variation dur system apps   robust variation dur strategy games   daily mean no. phone ringing
ES4, self-consciousness	Nightly mean no. calls   daily mean no. phone ringing   daily mean no. contacts calls   daily mean no. outgoing calls   daily mean no. contacts incoming calls

The top five most predictive features are shown for each successfully predicted personality dimension in the random forest models. The ranking is based on permutation feature importance and goes from left (high) to right (low). dur = duration.

Source: Stachl et al (2020)

- In this work (Stachl et al, 2020), a total of 743 volunteers were recruited via Forums, social media, Blackboards, Flyers, and Direct-recruitment, between Sept 2014 and Jan 2018. This focuses on Big 5 personality dimension identification (Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability), as shown in Figure 2.18.

- Random Forest method is incorporated.
- Behavior in Domains of:
  - Communication and social behavior
  - Music composition
  - App Usage
- Mobility
  - Overall phone activity
  - Data and night time activity

### 2.5.6 Predicting Negative Emotions based on Mobile Phone Usage Patterns: An Explanatory Study

Figure 2.19  
Predicting Negative Emotion: Emotion scales and Feature selection



Figure 3. Visual analogue scale for anxiety.

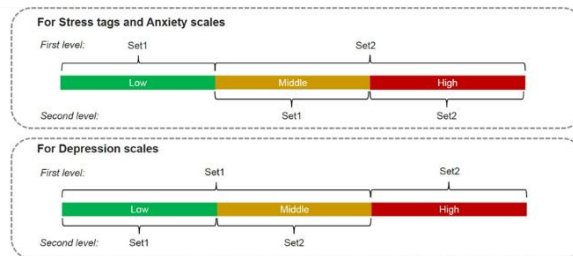


Figure 4. Two-level feature selection for scales of negative emotions.

Source: Hung et al (2015)

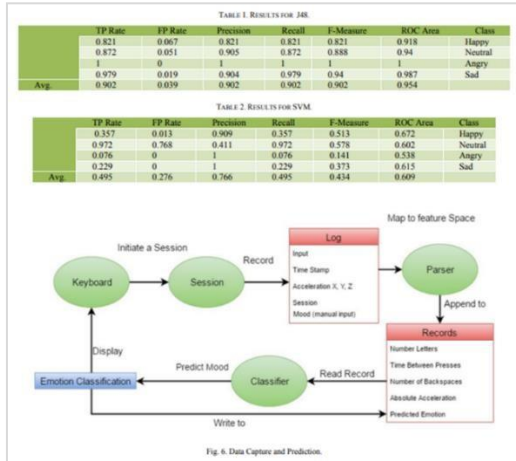
- Predicting 3 negative emotions: Depression, Anxiety and Stress (3 levels for each: low, medium, high, where low is neutral/no negative emotions), as shown in Figure 2.19 (Hung et al, 2015)
- Dataset: Self collected, Phone call usage, duration, No. of times apps opened, type of apps, average gap between 2 calls (this was collected for 3 hours window)
- 28 participants for Training and 18 participants for Evaluation (14 days data for training and 5 days data for Evaluation)
- Feature selection – 4 types of techniques are tried, Model – Support Vector Machines (SVM), Naïve Bayes (NB), Decision Tree
- Time slot for Testing – 0.5, 1, 2, 3 hours → Best 2 hours of data collection and prediction
- Data collection → Developed own application to log the data
- Model trained offline and deployed

### **2.5.7 Emotion Recognition using Mobile phones**

- The system uses accelerometers readings and various aspects of typing behavior like speed, number of backspaces and time-delay between letters to train a classifier to predict emotions
- Dataset: Self collected, Keyboard usage, accelerometer, time-delay between letters, backspace usage count
- Naïve Bayes (NB), Decision Tree (in specific, J48 Random Tree algorithm), Instance Based Learner (IBk classifier), Multi-response Linear Regression and SVM were evaluated, and J48 was found to be the best classifier with over 90% accuracy and precision.

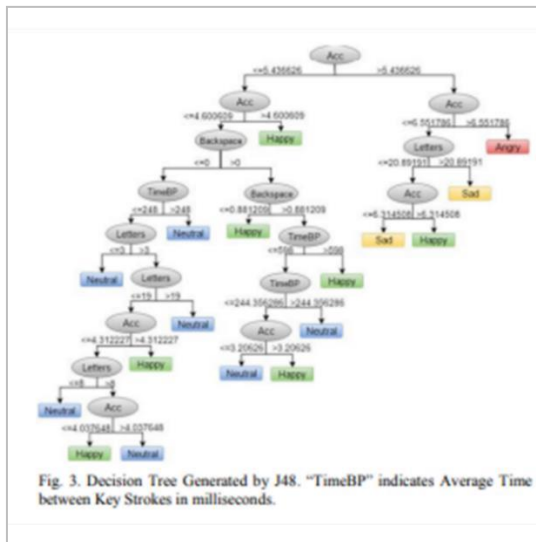
- The model architecture is captured in Figure 2.21 and the data pipeline, prediction and results are captured in Figure 2.20 (Shapsough et al, 2016)

Figure 2.20  
Emotion Recognition: Data Capture, Prediction and Results



Source: Shapsough et al (2016)

Figure 2.21  
Emotion Recognition: Decision Tree (J48), System Architecture and Experience



Source: Shapsough et al (2016)

## **2.6 Summary**

Most of these prior works rely on the content of user activity, like message content which may contain user-private information. Further, most studies collect data by incentivizing the user to record emotions which may result in unnatural and forced data. We address these drawbacks in our work and propose a non-intrusive smartphone activity annotation technique followed by some significant insights.

CHAPTER III:  
RESEARCH MOTIVATION & OBJECTIVES

**3.1 Opportunity Areas**

1. **Multimodal** Emotion Recognition: Even though there are few DNN architectures explored, this area is not matured. The existing prior arts are not fully viable for deployment and commercialization of real-time application.

2. **Demographic** and User Activities for Emotion Recognition: The prior arts discuss utilizing data that belong to one or more modalities, for emotion recognition. However, users perform various activities in their smart devices (ex: using various applications and performing various tasks in smart phone). Further, user profile and demographic information like age, gender, location, etc. are unexplored for emotion understanding. This is one of novel areas to be explored, to bring intuitive emotion-based insights. This requires utilizing both structured and unstructured data, and will involve a unique machine learning model architecture.

3. **Wearable** Data for Emotion Recognition: The existing architectures talk about individual-sensor based emotion detection. However, utilizing all the data in wearable for emotion understanding is still an area to be experimented. For example, various sensors data and their usage analytics available in Smart Watch, activities performed in Smart Watch and so on.

4. **Cross-device** Emotion Understanding: If the demographic data discussed in point 2 above, not only in user's smart phone, but also in user's other devices like Tablet, Books etc. are combined, along with Wearable data explained in point 3 above, then, this



will be an innovative and state-of-the-art solution, that will be introduced first time in the market.

5. **Other** Opportunity areas in Emotion Recognition: Privacy-preservation in Emotion recognition tasks, in the above-mentioned problems, based on known and unknown type dataset, is an unsolved research problem. Further, utilizing the above-mentioned opportunity areas, for identifying and coming up with intuitive business applications is also a challenging research task.

### **3.2 Discussion**

There are multiple opportunity areas mentioned in the previous section. To simplify, the main objective of the research is to come up with novel deep learning model architecture which embeds categorical and unstructured information. The goal of this project is also to ensure that the proposed model is common for any classification task (beyond emotion), which constitutes similar datasets having categorical data (ex: user demographics, wearable, cross-device datasets) and unstructured data (ex: text, image, audio, video datasets). Finally, the research task involves coming up with end -to-end pipeline for deploying emotion recognition model.

### **3.3 Research Questions**

1. What are the State-of-the-art model architectures, that are feasible for commercialization and real-time applications, and that can be inferred with low computation devices (like smart phones)?

2. How can the data be collected - demographic data, user profile, user's device state, user activities in the devices?

3. How can we mitigate privacy issues while collecting above mentioned data from the user and while processing user data for emotion recognition tasks?

4. What are the insights from the collected data, before performing emotion recognition task?

5. What should be the model architecture for categorical data collected for emotion recognition and how should the end-to-end pipeline look like? (Data -> Novel Fusion Embedding -> unique fusion-model-architecture -> KPI evaluation for real-time applications (Inference) -> Beta Deployment).

6. What is the current state-of-the-art (SOTA) KPI and what are the optimal KPIs that should be considered for the model?

### **3.4 Selected Problem Statements & Objectives**

Considering all the above factors, the main contributions of this research work are manifold:

1. **Emotion Data collection Experience:** Prepare methods for data collection, and design emotion data collection experience. [Research Question 2]

2. **System Design and Data collection:** Propose a novel annotation technique using the smartphone for generating ground truth labels. Develop a user trial app based on the same and distribute it to the participants for collecting tagged ground truth data. Identify personally identifiable information (PII) to protect privacy and incorporate it while collecting data. [Research Question 2 and 3]

3. **Emotion Data analysis:** Derive critical insights from the data collected from the user trial app about emotion, personality, and user behavior. [Research Question 4]

4. **Feature Engineering & Emotion Prediction:** Design a model having the relation between personality, emotion, and user smartphone behavior in a non -intrusive way, maintaining user privacy. [Research Question 5]

5. **Result and Analysis:** Validate the collected non-intrusive data by developing and implementing a system for automatic emotion detection, extract different features from the collected tagged data, train the machine learning model, test its performance and compare with SOTA KPIs. [Research Question 6]

### 3.5 Data Collection Objectives

- **WHAT:** Categorical features (app launch, activity, etc. – will be detailed in next section), tagged with emotions
- **WHO:** User trial (Participants)
- **WHERE:** Smartphone (smart watch in future)
- **WHEN:** Periodic and Direct
- **WHY:** To establish ground truth
- **HOW:** 1. Direct feedback, 2. Non-intrusive (Data collection application)

Important points for Data collection:

- No Personally Identifiable Information (PII) will be collected
- No 3<sup>rd</sup> party application data will be parsed or extracted
- SMS contents will not be not read.
- Adequate measure will be taken so that data cannot be traced back to originator (Encryption, Anonymization, Data –Deid)
- This data collection exercise is required to build a predictive model for User Emotion Understanding

- The data collected split into train and test set and used to build a machine learning model. The data will not be used for marketing or any other purpose.
- The application will be distributed for data collection.

### 3.6 Categorical Features Selection

*Table 3.1  
List of Categorical Features*

S. No.	Parameter	Research Paper	Android API	Features
1	Calls	<ol style="list-style-type: none"> <li>1. Predicting Personality from patterns of behavior collected with smartphones</li> <li>2. iSelf: Towards cold-start emotion labelling using Transfer Learning with smart phones</li> <li>3. MoodScope: Building a mood sensor from smart phone usage patterns</li> <li>4. Daily stress recognition from mobile phone data, weather conditions and individual traits</li> <li>5. Predicting negative emotions based on mobile phone usage patterns: An explanatory study</li> </ol>	Yes	<ol style="list-style-type: none"> <li>1. No of calls made</li> <li>2. Duration of calls</li> <li>3. Missed calls</li> <li>4. Time elapsed between consecutive calls</li> <li>5. No of calls received</li> <li>6. VOIP calls made</li> </ol>

2	Contact entries	<ol style="list-style-type: none"> <li>1. Predicting personality from patterns of behavior collected with smartphones</li> <li>2. iSelf: Towards cold-start emotion labelling using Transfer Learning with smart phones</li> <li>3. MoodScope: Building a mood sensor from smart phone usage patterns</li> <li>4. Daily stress recognition from mobile phone data, weather conditions and individual traits</li> <li>5. Predicting negative emotions based on mobile phone usage patterns: An explanatory study</li> </ol>	Yes	<ol style="list-style-type: none"> <li>1. Hashed contact count based on frequency – Top 10 / 15</li> <li>2. Duration – Top 10 / 15</li> </ol>
3	Texting	<ol style="list-style-type: none"> <li>1. Predicting Personality from patterns of behavior collected with smartphones</li> <li>2. iSelf: Towards cold-start emotion labelling using Transfer Learning with smart phones</li> <li>3. MoodScope: Building a mood sensor from smart phone usage patterns</li> </ol>	No	Ignore

		4. Daily stress recognition from mobile phone data, weather conditions and individual traits		
4	Global Positioning System (GPS), Location	<ol style="list-style-type: none"> <li>1. Predicting Personality from patterns of behavior collected with smart phones</li> <li>2. iSelf: Towards cold-start emotion labelling using Transfer Learning with smart phones</li> <li>3. MoodScope: Building a mood sensor from smart phone usage patterns</li> </ol>	Yes	<ol style="list-style-type: none"> <li>1. Latitude</li> <li>2. Longitude</li> <li>3. Place</li> </ol>
5	App starts / Installations	<ol style="list-style-type: none"> <li>1. Predicting Personality from patterns of behavior collected with smart phones</li> <li>2. iSelf: Towards cold-start emotion labelling using Transfer Learning with smart phones</li> <li>3. MoodScope: Building a mood sensor from smart phone usage patterns</li> <li>4. Does Smartphone Use Drive our Emotion or vice versa? A casual Analysis</li> </ol>	Yes	<ol style="list-style-type: none"> <li>1. App installation (package name)</li> <li>2. App Launch</li> <li>3. App Usage Duration</li> </ol>

6	Screen De/activation	1. Predicting Personality from patterns of behavior collected with smart phones	Yes	1. Screen Lock / Unlock Number 2. Time between consecutive locks / unlocks
7	Flight mode De/activation	1. Predicting Personality from patterns of behavior collected with smart phones	Yes	1. Current modes: Airplane, Silent, Normal
8	Calls	6. Predicting Personality from patterns of behavior collected with smart phones 7. iSelf: Towards cold-start emotion labelling using Transfer Learning with smart phones 8. MoodScope: Building a mood sensor from smart phone usage patterns	Yes	1. Bluetooth connection 2. Bluetooth connection to which it is connected 3. Duration 4. Surrounding Bluetooth connection IDs
9	Booting Events	1. Predicting Personality from patterns of behavior collected with smart phones	Listening broadcast	System Restart
10	Played Music	1. Predicting Personality from patterns of behavior collected with smart phones	Yes	Music played?

11	Battery Charging status	1. Predicting Personality from patterns of behavior collected with smart phones	Yes	Battery status
12	Photo and Video Events	1. Predicting Personality from patterns of behavior collected with smart phones	NA	1. Image capture Events 2. Video capture events 3. Mic recording events
13	Connection to wireless networks (Wi-Fi)	1. Predicting Personality from patterns of behavior collected with smart phones 2. iSelf: Towards cold-start emotion labelling using Transfer Learning with smart phones	Yes	1. Wi-Fi status 2. No of Wi-Fi signals in each time slot
14	Character length of text messages	1. Predicting Personality from patterns of behavior collected with smart phones 2. iSelf: Towards cold-start emotion labelling using Transfer Learning with smart phones 3. MoodScope: Building a mood sensor from smart phone usage patterns	NA	Ignore
15	Technical device	1. Predicting Personality from	NA	One-time



	characteristics where collected	patterns of behavior collected with smart phones		
16	Irreversibly hash-encoded version of contacts and phone numbers were collected to enable us to measure the number of distinct contacts while preventing the possibility of reidentification	1. Predicting Personality from patterns of behavior collected with smart phones	Repeated	This logic will be used while collecting data of contact entries
17	SMS Contacts	1. iSelf: Towards cold-start emotion labelling using Transfer Learning with smart phones	Yes	1. Frequency of SMS received
18	On-line Social networking site contents (Keywords, Tags, etc.)	1. iSelf: Towards cold-start emotion labelling using Transfer Learning with smart phones	NA	Not possible
19	Browser context (Bookmark, Search content)	1. iSelf: Towards cold-start emotion labelling using Transfer Learning with smart phones 2. MoodScope: Building a mood sensor from smart phone usage patterns	Yes	Browser Contents which we can get (to explore)

20	Sensor	<ol style="list-style-type: none"> <li>1. iSelf: Towards cold-start emotion labelling using Transfer Learning with smart phones</li> <li>2. Emotion Recognition using Mobile phones</li> </ol>	Yes	<ol style="list-style-type: none"> <li>1. Accelerometer Data</li> <li>2. Gyroscope Data</li> </ol>
21	No. of contact associated with missed calls	<ol style="list-style-type: none"> <li>1. MoodScope: Building a mood sensor from smart phone usage patterns</li> </ol>	Repeated	Covered
22	Entropy of call, contacts and SMS (Randomness)	<ol style="list-style-type: none"> <li>1. MoodScope: Building a mood sensor from smart phone usage patterns</li> </ol>	Yes	To explore
23	Call, Contacts to Interaction Ratio	<ol style="list-style-type: none"> <li>1. MoodScope: Building a mood sensor from smart phone usage patterns</li> </ol>	Yes	Out of saved contacts, how many are we contacting.
24	Percent Call during Night	<ol style="list-style-type: none"> <li>1. MoodScope: Building a mood sensor from smart phone usage patterns</li> </ol>	Repeated	Night calls (10 PM for ex.)
25	SMS Response Rate	<ol style="list-style-type: none"> <li>1. MoodScope: Building a mood sensor from smart phone usage patterns</li> </ol>	NA	Can't collect
26	SMS Response Latency	<ol style="list-style-type: none"> <li>1. MoodScope: Building a mood sensor from smart</li> </ol>	NA	Can't collect

		phone usage patterns		
27	Percent SMS and Calls Initiated	1. MoodScope: Building a mood sensor from smart phone usage patterns	NA	Not required
28	Time elapsed between 2 events (Call, SMS) – Average and Variance	1. MoodScope: Building a mood sensor from smart phone usage patterns	No direct API	App switch frequency
29	Facial Expressions	1. Does Smartphone Use Drive our Emotion or vice versa? A casual Analysis	NA	Not possible
30	Calling states – Idle, Ring, In-call, Duration, Gap between calls	1. Predicting negative emotions based on mobile phone usage patterns: An explanatory study	Covered	Covered
31	Average Time delay between typed letters	1. Emotion Recognition using Mobile phones	NA	Required Keyboard service level changes
32	No of backspaces	1. Emotion Recognition using Mobile phones	NA	Required Keyboard service level changes
33	Calendar Events	NA - Based on smart phone usage	Yes	1. Events duration 2. Events Title

34	Sharing	NA - Based on smart phone usage	Yes	No. of times user has shared any content
35	Notification	NA - Based on smart phone usage	Not available	1. No of notifications received 2. No of notifications Interacted
36	Lock screen / Launcher	NA - Based on smart phone usage	Yes	To explore
37	[Extras] Detect change in user activity like walking, running, etc.	NA - Based on smart phone usage	Yes	To explore

### 3.7 Data Collection: Risk and Mitigation

1. Data collected through data collection app may not provide enough direction.  
To mitigate, weekly dump of data needs to be verified for consistency.
2. Not enough data samples for correct predictions.  
To mitigate, after size-able data is collected, pattern will be identified from already existing (collected) data after the ground-truth is ascertained.

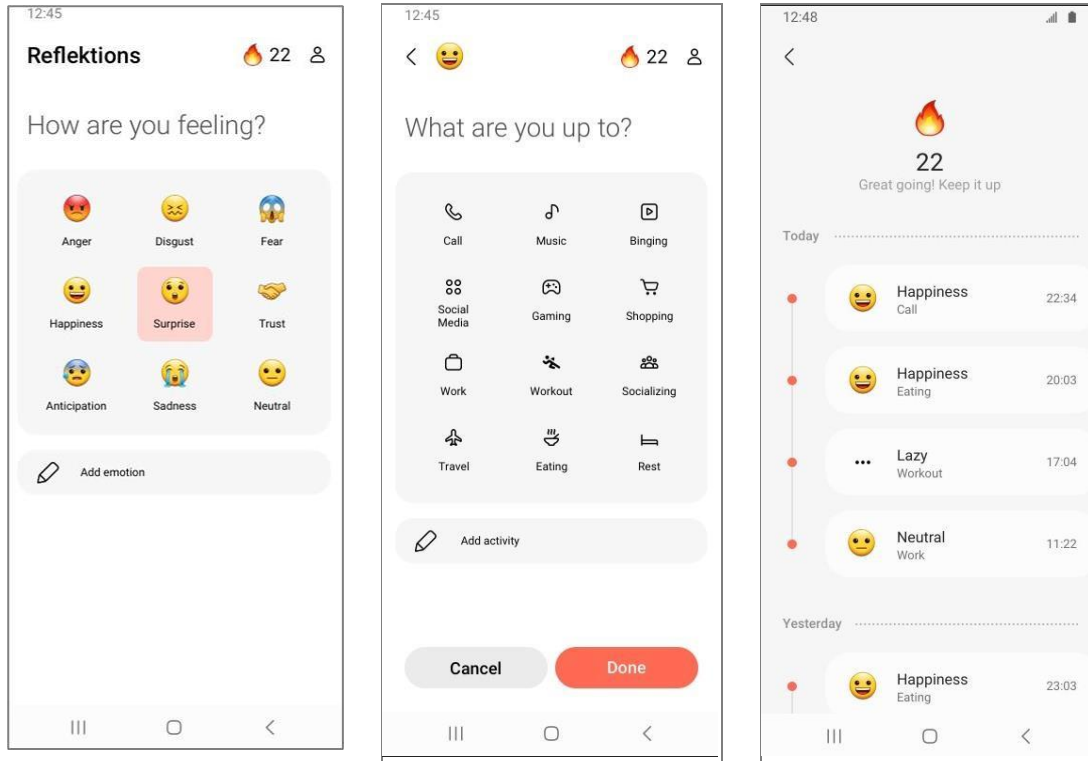
## CHAPTER IV: METHODOLOGY

### **4.1 Designing Emotion Data Collection Experience**

Designing the most effective application for data collection is a challenge. The user needs to be motivated to regularly contribute annotations without the need to provide any incentive for the same. Doing this will help collect accurate data annotation and avoid forced data logging. To achieve this, a novel design is developed for data collection, the details of which are discussed in this section.

The emotion annotation method is being designed following an iterative and user-centric approach. The experience is designed not only to enable users to record emotion-related data with ease but also to reflect on the data recorded insightfully. To encounter the main problem, which any data recording app might face (Caldeira et al., 2017), which is to motivate or trigger users to record the data every day in a regular fashion. This will be solved by studying social media behaviors where users feel rewarded subconsciously for using the app, e.g., Gamification in Snapchat (Hamari et al., 2014). They introduced a feature called Streaks that encourages users to open the app and use it regularly (Furner et al., 2014). Taking from the same and incorporated Streaks as an inspiration in our design to motivate users for frequent data annotation as shown in Figure 4.1c.

*Figure 4.1*  
*Main Screens of the Reflektion Application*



a. Emotion selection screen    b. Activity selection screen    c. Streak page

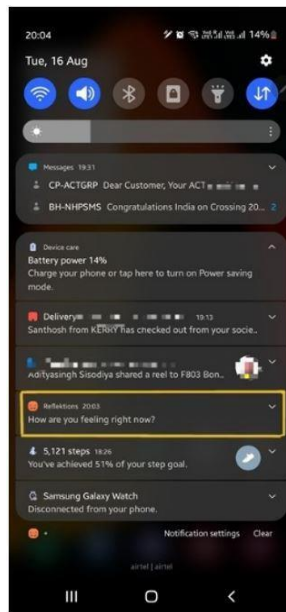
Further, it is anticipated that users might find it challenging to find the app repeatedly, as they might not have placed it in an accessible manner on the phone (Lavid Ben Lulu et al., 2013). Thus, a nudge is introduced, which prompts the user to add the app widget on the home screen for quick access. This nudge surfaces during the first-time experience of the app (Figure 4.2c). Once the user adds the widget to the home screen, they can tap on it to open the emotion selection step to record their emotion quickly.

Figure 4.2

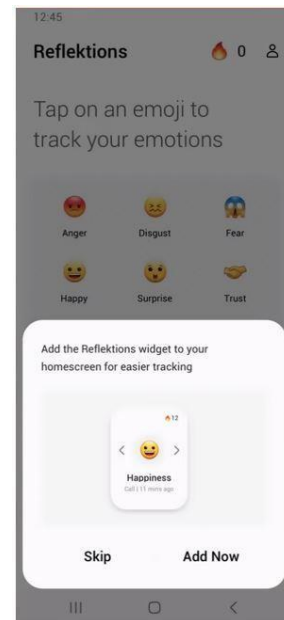
Periodic notifications for data collection & Widget nudge for Data annotation



a. Silent Notifications



b. Revisiting Notifications



c. Widget Nudge

The periodic notifications are added, which serve as a reminder for the user to log their emotion. Apart from these nudges, we also trigger notifications after certain activities like a switch in airplane mode, phone reboot, playing music, etc., which allow us to collect data in that particular context. Further, the users can then reflect on the emotion they viewed just after the event/ situation has ended (Ferdinando et al., 2018).

The primary consideration for app design was to make it compatible and usable for all android devices devoid of screen size and resolution. Hence, some of these rules are followed, for basic app design laid out by User experience and android design guidelines (R Fulcher, 2022). The language used in the app is straightforward and does not require users to spend a lot of time comprehending what is expected of them (Chloe Blanchard,

2022). E.g., Emotion selection page – “How are you feeling?”, “Tap on an emoji to track your emotions,” etc. Activity selection page – “What are you up to?” etc.

Users use emoticons to communicate their emotions on digital platforms (Yus et al., 2014). Thus, the emotions are symbolized using emoticons for quick learnability and repeated accurate input of the emotions they were feeling. The emoticons are used not according to the emotion they originally represented, but that match the emotion to be recorded as per the common perception among most users, e.g., Gen Z, Millennials, etc. (Hoefel F Francis T, 2022). Activities that insight emotion in users can be bound to users’ smartphones or physical activity outside the phone. Therefore, the first two rows of icons of subsequent activity screen represent activities on the phone, and the next two are physical activities outside the smartphone.

Before the participants can start providing emotion data, our data collection Reflektion app requires one time on-boarding setup. During on-boarding, participants are asked to provide basic demographic data (age-range, gender and occupation status) and big five personality traits (extraversion, openness, conscientiousness, agreeableness, and neuroticism) as discussed in section 1. Each trait is provided with three descriptors as low, medium and high which represent level of each personality trait. To ensure most appropriate selection for personality traits by participants, app UI provides description and expected behaviour for each trait. Users can then easily follow the simple three -step process to reach the emotion selection page through various nudges, select emotion, and select related activity to record the emotion data. Its initially started with nine emotion categories and further expanded to three more emotion categories based on user feedback in subsequent release of the application. In the future, the aim is to provide insights based



on the patterns emerging from the emotion data on the application screen. Until then, users can visit the streak page to have an overview of the emotion data they recorded for self-reflection on their day, week, month, and more.

## **4.2 System Design and Data Collection**

Post designing the user interface and application structure, implementing an effective system functionality is another challenge. We take inspiration from LiKamWa et al (2013) and attempt to further eliminate some gaps. We observe a need to collect the smartphone activity data in the background before and after the user records their emotion to capture the details of what could have led to the emotion and the user behavior in a particular emotional state. Our data collection window lasts 45 minutes, and we trigger a notification 15 minutes into the collection window (data collected for 15 minutes before user-annotation, 30 minutes post the annotation). For user convenience, we prompt the user to annotate the data with a frequency of 4 times daily, between 9am and 9 pm. This ensures that participants do not disable the app, fearing battery drain. However, the participants can enter emotion anytime by directly launching the application. Participants were made aware of the data and data collection process to ensure transparency. We design the data collection, and annotation application with these policies after detailed user-trial experimentation and research (LiKamWa et al., 2013). For deciding the data collection window, we take cues from previous studies by MoodExplorer (Zhang et al., 2017) and EmoSensing (Roshanaei et al., 2017) and experimented with one-hour window size during the early phase of data collection. However, we observe that multiple emotions are being reported by participants due to the large window size, leading to undesired data collection. So, based on user feedback and our data analysis, we reduced the data collection window to 45 minutes.

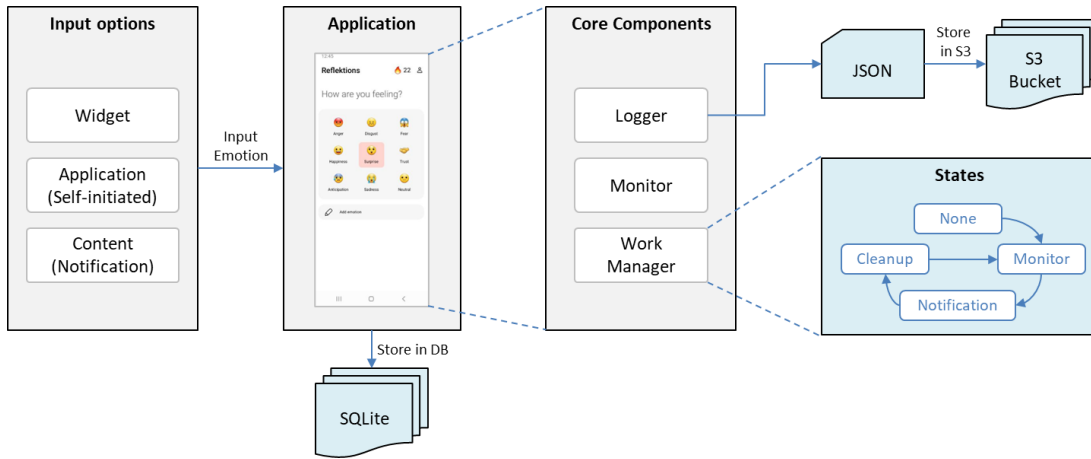
The designed application allows the user to record emotion in multiple ways: Through Scheduled Notifications, the widget, and self -initiated, as explained in the previous section. The application system design has mainly three core components shown in the Figure 4.3: 1) Work Manager, 2) Monitors, and 3) Logger.

The Work Manager is responsible for scheduling tasks periodically by transitioning among different states. These states are None, Monitor, Notification, and Cleanup. The application enters the None state immediately after installation, then enters the Monitor state where we register for required Monitors to capture data. We then register for the Notification state with a 15-minute delay. In this case, we give the user a gentle nudge to enter their current emotion, with an optional activity entry and a short note that allows us to capture the trigger for the entered emotion. The monitors continue to function while in the notification state. The work manager registers for the cleanup state with a 30 -minute delay after the notification, where we unregister all monitors and retrieve the collected data from the database to convert it to JavaScript Object Notation (JSON) files. The Monitor state is registered with 120-minute delay. The Logger then attempts to log the collected data into an Amazon Web Service (AWS) Simple Storage Service (S3) bucket that holds the data accumulated from all the users. We determined the duration of these delays through extensive user-trial experimentation.

The Monitors are various broadcast receivers and listeners that help capture significant user activities like App usage time, Wi-Fi connection/disconnection, Flight Mode On/Off, etc. Some features like Call duration are further used to derive essential elements like average call duration for the top 10 most contacted individuals. These are

non-intrusive features collected through user consent. We ensure to mask any PII if present in any collected user data.

Figure 4.3  
System Design



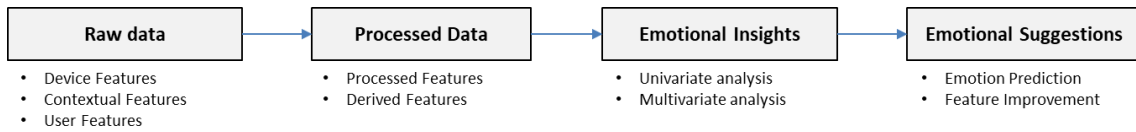
The Logger takes care of sending the user data to the AWS S3 Bucket. This first checks for any working network connection. If the device is connected, it checks for any pending JSON files for logging. Once we determine that there are indeed files pending, the Logger fetches the JSON files and attempts to log into the server. Once successful, the files are deleted from internal memory. Note that we use the secure android identifier to uniquely identify each device and avoid any personal information from being acquired. The files corresponding to respective devices are stored in separate folders.

### 4.3 Emotional Data Analysis

One of the crucial components of the entire process is data analysis. Analytics is used to process the collected raw data and further turn it into insights (that can unearth new relationships between features). Additionally, we can use these insights to provide

actionable recommendations and feature improvement suggestions as shown in Figure 4.4. This section provides an example of different analytics performed on the collected data, to demonstrate its potential.

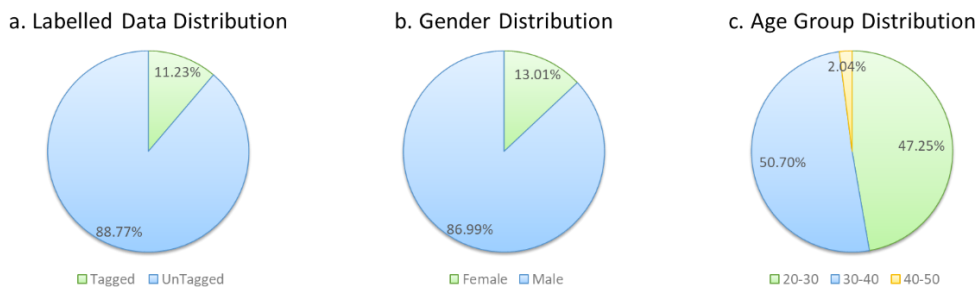
*Figure 4.4  
Data Processing Pipeline*



### 4.3.1 Research Context

The android-based annotation application called Reflektion is used to collect the data. The annotated data is collected from November 2022 to April 2023. As mentioned in previous sections, this study aims to understand user conduct, emotion, and personality and explore relationships that can enable novel use cases. The participants used Android-based smartphones having Android Pie and above version. The participants are given the freedom to log the emotion data or ignore the prompt, thereby eliminating any need for forced entry to strengthen the authenticity of the data further.

*Figure 4.5  
Ground Truth Labels Tagging & Demographic Data Distribution*



As shown in Figure 4.5a, approximately 11 percent of the data collected is tagged by the user with emotion. 13% of participants are female and 87% male. A major percentage (98%) of participants are within the age group of 20 -40 years. This demographic distribution is covered in Figure 4.5b and 4.5c respectively. This dataset is not entirely representative of the user’s smartphone usage behavior, mainly due to demographic constraints (all the participants are from a specific country). Also, all the participants are employed. Nevertheless, the outcomes highlight the approach’s unique value.

### 4.3.2 Data Visualization and Emotional Insights

Under this, we plot individual features and try to derive insights from the same. One crucial finding is that the existing eight basic emotions are not enough to capture user emotions non-intrusively. Instead, few other emotion categories get priority over the different categories of eight basic emotions, as shown in Figure 4.6a. Figure 4.6b shows the distribution of other emotions manually entered by the participants.

Figure 4.6  
Emotion Distribution labelled by the Users

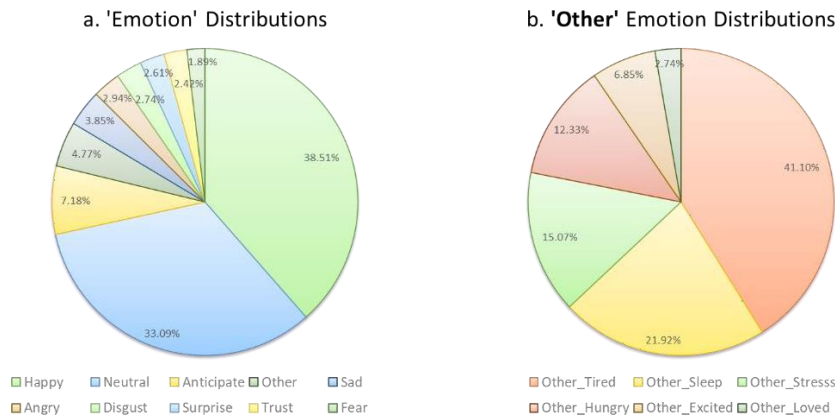
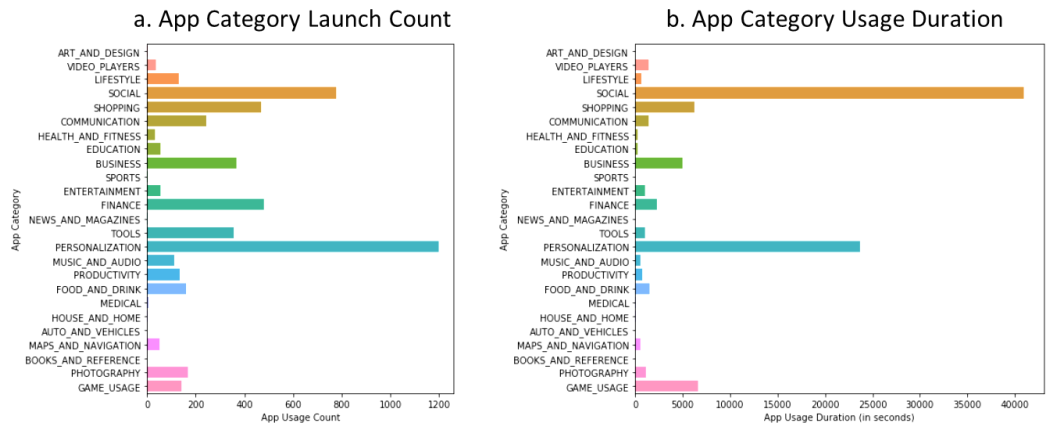


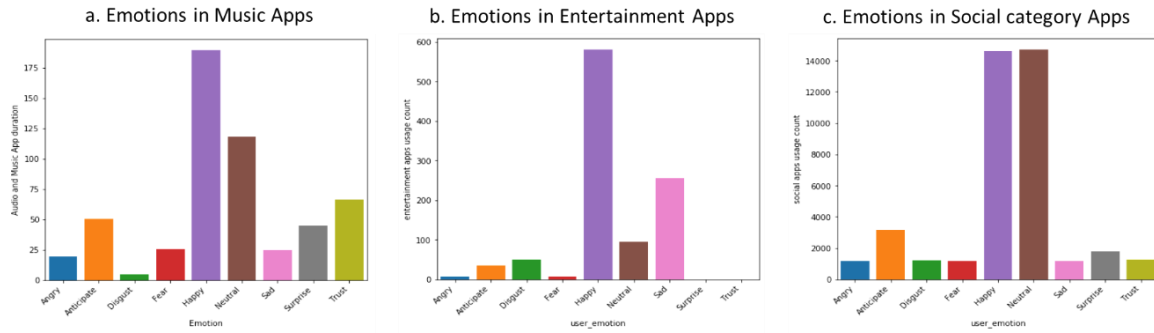
Figure 4.7 captures the app category launch and duration. Though personalization apps were launched more often than social apps, social applications' usage duration is comparatively higher. Shopping and Game are the other two app categories where the user spends most of the time.

*Figure 4.7*  
*Distribution of application categories (Launch count & Usage duration)*



The distribution of significant application categories usage for various emotions is covered in Figure 4.8. A social category app is used mainly by the user in a happy or neutral emotional state. One fascinating insight is that user tends to use entertainment applications when they are sad. When the user is disgusted, music-related applications are rarely used. Apart from happy and neutral emotions, when the user's emotional state is that of trust as well, music applications are used.

*Figure 4.8*  
*Emotion Distribution in different category of applications (App Usage vs Emotion)*



Users tend to lock/unlock phones more when in an anticipation state as shown in Figure 4.9a. In anticipation state, apart from high lock/unlock, foreground time (refer Figure 4.9c) and food/drink category app usage is also high compared to other emotional states. This can be when user orders food, keeps lock/unlock the phone and enters into anticipation state. Figure 4.9a shows lesser lock/unlock frequency for sad emotion state and Figure 4.9c also shows comparatively low values for foreground time for sad emotion. From these observations, it can be weakly inferred that user tends to use phone less when in sad state. It's also observed that users tend to be in angry state when missed call count is high (refer Figure 4.9d). High incoming calls are associated with anticipation state can be seen in Figure 4.9b. For all the graphs in Figure 4.9, X-axis represents count/usage (in minutes) of feature fitted in fixed size bins, Y-Axis represents number of Data points available in the bins.

*Figure 4.9*  
*Emotion Distribution among various smart phone features*

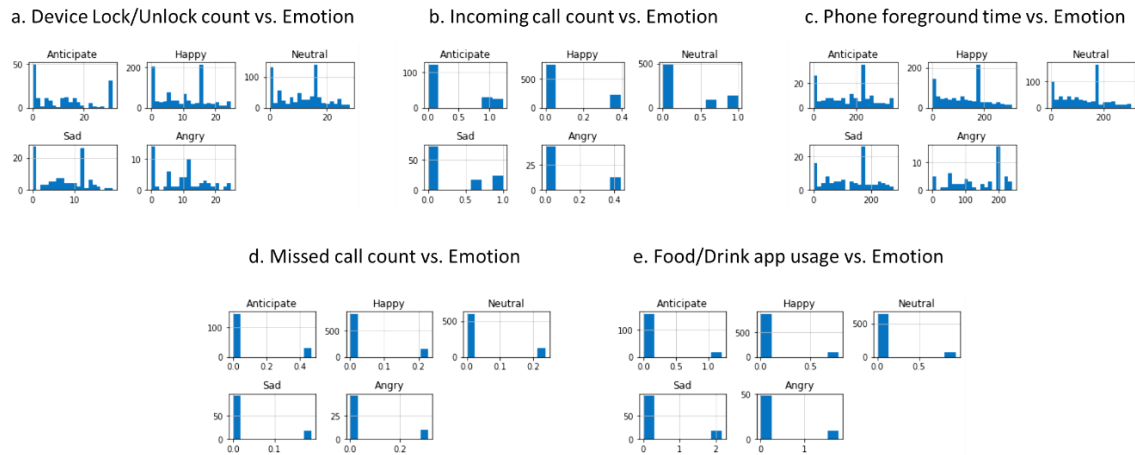
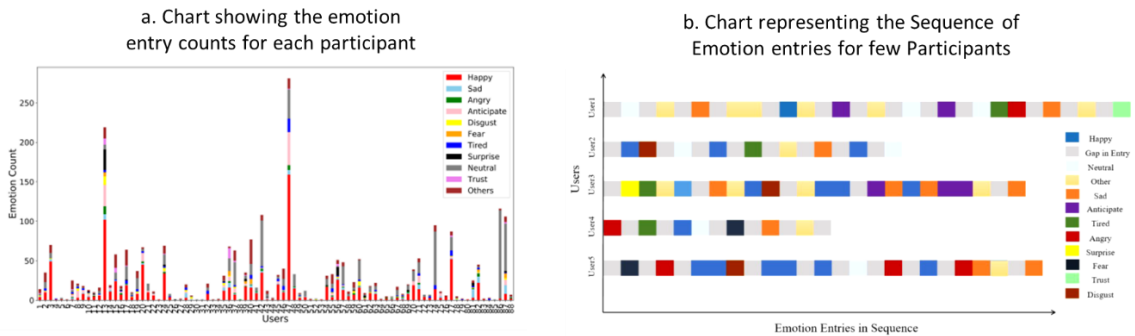


Figure 4.10 captures emotion dynamics of the emotion data logged by the user. Figure 4.10a captures the count of different emotions entered by individual users over the data collection period. Figure 4.10b captures the distribution of emotion entered by various users in sequential manner affirming the fact that user undergoes through various emotions over a period of time and our data captures the variation.

*Figure 4.10*  
*User Emotion Dynamics*



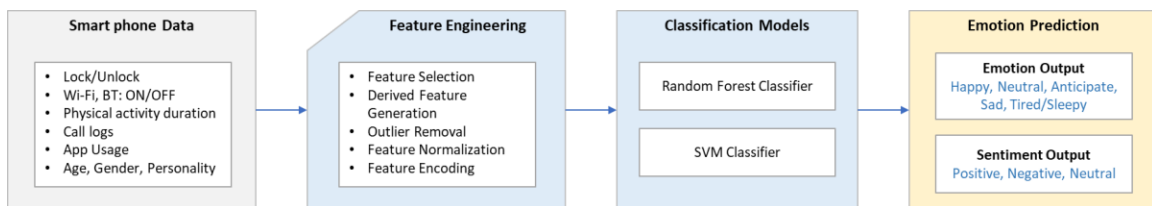


#### 4.4 Feature Engineering

As a first step, we collect user data through android based Reflektion App as mentioned in section 4. Further, this data is encrypted and sent to the server in JSON format. JSON data files are downloaded and further processed to convert into Comma Separated Values (CSV) format in the next step to draw actionable insights and model development. We use python-based environment for all our data pre-processing and model development.

Figure 4.11 describes our proposed model pipeline. We use heat map-based feature correlation analysis to understand the correlation between features. We identify that 'application added' and 'application replaced' are highly correlated and hence dropped the 'application replaced' feature. We also observe that some of the captured features always have null values such as physical activity corresponding to foot and tilting, and application-changed events, so we decided to drop these.

*Figure 4.11  
Overview of Model Prediction Pipeline, with Novel smartphone data*



The following are the features from the data that are collected in the beta application.

- Bluetooth (On/Off, Pairing/unpairing, Device discovery, Device connected/disconnected)
- Contacts (Missed call, incoming/outgoing call: count, duration)
- App state change (install/ uninstall/update/replaced)
- Lockscreen (lock/unlock)
- Wifi (Scan results, on/off, network connected/disconnected)
- SMS (Status)
- Physical activity (vehicle in, walking, running, on bicycle, still)
- Phone Mode (Flight Mode, Silent, Vibrate, Normal)
- Music (Metadata changed)
- App Usage (Top 10 app packages used, duration): Once per window
- Calendar (Event info, event time): Once per window
- System Parameters (Battery percentage, RAM, ROM)

*Table 4.1  
Application Categories*

Category	Sample Applications
Social	Twitter, Reddit, Instagram
Entertainment	MyGalaxy, Netflix, SonyLiv
Finance	Phonepe, Paytm, Gnet, ICICI-iMobile, Splitwise
Food and Drinks	Zomato, Swiggy, Dineout, Grofers
Music and Audio	Youtube Music, Spotify, Dolby On
Personalizing	Microsoft Launcher, Nova Launcher
Lifestyle	Amazon Alexa, NoBrokerHood, Google Home
Communication	WhatsApp, Telegram, Contacts, Messaging
Health and Fitness	Cure.Fit, Fitbit, Tranquil-Silver Oak Health, Aarogyasethu
Business	Microsoft Teams, LinkedIn, Glassdoor, Zoom
Shopping	Nykaa, Zepto, Lenskart, Olx, Bigbasket

The categories defined in Google Play is used to group the applications into 25 different categories as demonstrated in Figure 4.7. The Table 4.1 also shows major categories and names of the applications that are grouped in them.

As a part of the pre-processing pipeline, we first remove outliers from non-categorical features and then apply standard scaler normalization on these features. For identifying outliers, we use the inter quantile range (IQR) to identify and replace outliers by mean values of the respective features. In our experiment, replacing outliers by mean instead of Q3 (third quantile) gives better results. We then encode categorical features such as age group, user activity, time of the day, five personalities (openness, agreeable, conscientiousness, extraversion, neuroticism), and phone mode using one hot encoding method and use label encoding for our target variable 'user emotion' and gender. For getting the time of the day, the entire day is divided into four segments: Morning (6-12), Noon (12-16), Evening (16-20), and Night (20-6) which makes it a categorical feature for our analysis. After the feature encoding step, we get a total of 59 features including the target variable which we use for model training and validation. As discussed in section 5.2, we get several emotion labels tagged by users but for the emotion prediction task, we drop those emotions whose frequency is less than 2% across the data. Consequently, we get five emotions viz Happy, Neutral, Anticipate, Tired/Sleepy, and Sad. We drop other emotions as the number of data points is not sufficient for training and validation. As shown in Data Visualization and Insights, the dataset is highly imbalanced with emotions such as happy and neutral having a higher proportion of samples compared to the other emotions. So, to handle the data imbalance issue, we try different standard sampling approaches such as Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2002) and SMOTE-based over and under-sampling strategies as well as class weights for better generalization

of the model on rarer classes. In our experiments, we get better results using class weights compared to other methods. So, for all experiments, we adopt class weights as a standard method for handling class imbalance.

Apart from emotion prediction, we also provide predictions for three sentiments polarity viz. positive, negative, and neutral on the same dataset. To generate sentiment ground truth for the data points, we use a study by Gonçalves et al (2017) to group different emotions into sentiments. Unlike for emotion task where emotions having frequency lesser than 2% are dropped, we consider all emotions for getting sentiment labels except the surprise and anticipation emotions as they can be both positive and negative based on the user activity outcome.

#### **4.5 Emotion Model – Model Experiments & Preliminary Results**

##### **Problem:**

- Accurately identifying emotion is a complex research area that can help to build multi-device experiences. Current research focuses on comprehending user emotion primarily through user-private data.
- Another issue is of collecting precise emotion ground truth labels. Present annotation techniques rely either on self-reporting or recording on desktop applications, which is less natural given that smartphones have emerged as the dominant form of communication.

**Solution:**

- Devising a user-centric approach to collect and annotate user data in a non-intrusive way in smartphones.
- Derive insights from the annotated data comprising user behavior, emotion and personality.
- The annotated data should comprise only categorical features that do not include Personal Identifiable Information (PII), thus preserving user privacy
- Validate the annotated data by an emotion prediction model.

**4.5.1 Release stats – 15 Mar 2023 (Version 4.0, Model Version 1)**

Initially Graph Convolution Network (GCN) is tried. However, the results are not satisfactory. The details about GCN are captured in the next section. The Random Forest (RF) is used to predict Emotion from Categorical features. The results are as follows.

*Table 4.2  
Emotion Model Results (Precision, Recall, F1-Score): Mar 2023*

<b>Emotion</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
Happy	80.35	95.07	87.09
Anticipate	93.75	57.69	71.42
Neutral	79.56	85.71	82.52
Sad	100	5	9.52
Tired/Sleepy	67.3	100	80.454
Other	0	0	0

- Total accuracy: 79.57%
- Dataset duration: From July 2022 to 5<sup>th</sup> Oct 2022

- Unique users: 106
- Total dataset points: ~22K
- Labelled data points: ~2K (10%)

*Table 4.3*  
*Emotion Model - Confusion Matrix: Mar 2023*

	Happy	Aniticipate	Neutral	Sad	Tired/Sleepy	Other
Happy	405	0	17	0	4	0
Anticipate	7	15	4	0	0	0
Neutral	32	0	222	0	5	0
Sad	13	0	5	1	1	0
Tired/Sleepy	0	0	0	0	35	0
Other	47	1	31	0	7	0

(Y-axis: Actual Values, X-Axis: Predicted Values)

**Latest stats on user feedback for emotion predictions (Version 4.0):**

- Total valid (Ground truth + predictions are present): 852
- Total correct: 678
- Accuracy: 79.57%
- Total prediction entries (predictions present but no feedback from user): 6956

**Observations:**

- Happy and Neutral seem to be the most accurately predicted and most dominant emotions for users
- Tired/Sleepy emotion is predicted with 100% Recall rate

- Sad is the most miss-classified emotion (Need to identify feature specific to this emotion from patterns to improve performance)

**Next steps:**

- Aim to improve user response rate through brainstorming for better experience of the application
- Improve model further to reduce error rate.

#### **4.6 Personality Model – Model Experiments & Preliminary Results**

User's Personality can be categorized with 5 big personalities. The 5 big personality traits are as follows:

1. **Openness:** It emphasizes imagination and insight the most out of all five personality traits.
2. **Agreeable:** Personality trait manifesting itself in individual behavioral characteristics that are perceived as kind, sympathetic, co-operative, warm, frank and considerate
3. **Conscientiousness:** Personality trait of being careful, or diligent. Conscientiousness implies a desire to do a task well and to take obligations to others seriously.
4. **Extraversion:** Extraversion (or extroversion) is a personality trait characterized by excitability, sociability, talkativeness, assertiveness, and high amounts of emotional expressiveness.
5. **Neuroticism:** Neuroticism is a personality trait characterized by sadness, moodiness, and emotional instability.

#### 4.6.1 Personality - Release stats (Version 4.0 Model Version 1)

Table 4.4

Personality Model Results (Precision, Recall, F1-Score): Mar 2023

		Personality Traits				
		Extraversion	Agreeable	Openness	Conscientiousness	Neuroticism
Macro	Precision	0.73	0.77	0.77	0.7	0.76
Average	Recall	0.71	0.77	0.71	0.64	0.7
	F1-Score	0.72	0.77	0.73	0.67	0.73

- Overall accuracy: 77.4%
- Model Size: 494 KB
- Inference Time: 0.40 s

#### 4.6.2 Personality - User Feedback Results: Version 4.0 Model Version 1 (Till 16 Jun 2023)

- Total number of users: 40
- Total number of feedbacks: 2807
- Total number of feedbacks with ground truth: 1532
- Total number of users with ground truth: 32
- Accuracy Table for Personality Prediction (User Feedback): captured below.



*Table 4.5*  
*Personality Model Results (Accuracy): Jun 2023*

<b>Personality Traits</b>	<b>Accuracy</b>
Openness	27.02
Conscientiousness	18.67
Extraversion	22.11
Agreeable	51.39
Neuroticism	25.98

#### **4.6.3 Tabnet Architecture**

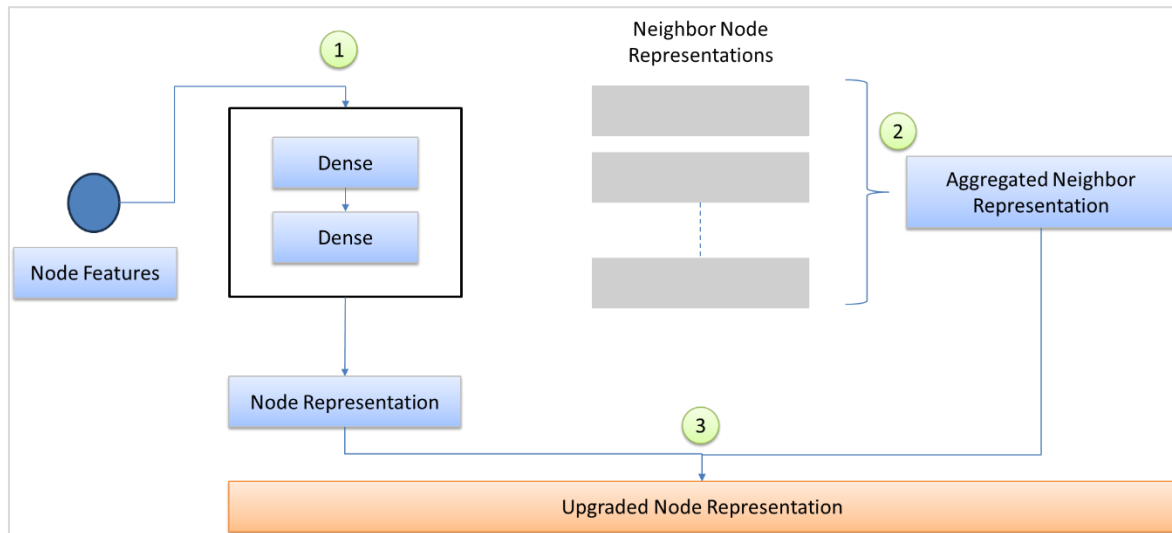
- Number of samples: 16895
- Number of features: 62

*Table 4.6*  
*Personality Model Results (Accuracy) – Tabnet Architecture: Jun 2023*

<b>Personality Traits</b>	<b>Accuracy</b>
Openness	70.875
Conscientiousness	72.79411
Extraversion	72.3278
Agreeable	72.184
Neuroticism	68.2747

## 4.7 Graphical Convolution Network (GCN) Experiment

Figure 4.12  
Graphical Convolution Network (GCN) Architecture: Experiment



Input:

- Edge Information:  $(2, \text{num\_edges})$  : Source and target info for each edge
- Edge weights:  $(\text{num\_edges})$
- Node Features:  $(\text{num\_nodes}, \text{feature\_size})$

Steps:

- Prepare: Node feature and basic dense network
- Aggregate: Aggregated Neighbor node representations
- Update: Upgrade the node representation

*Table 4.7*  
*GCN Results for Emotion Prediction*

<b>Model Version</b>	<b>Brief Description of Features</b>	<b>Train Accuracy</b>	<b>Test Accuracy</b>	<b>Size</b>
1	Simple base line model with mean aggregation and concat updation	33.33%	-	-
1.1	Custom dense aggregation and Gated Recurrent Unit (GRU) updation	40.38%	-	-
1.2	Custom dense aggregation and GRU updation with embeddings for Categorical Data	42.25%	46.67%	3.21 MB

This experiment is tried based on SOTA architecture study of utilizing Graph based network for Categorical features. However, as shown in the above table, the accuracy of model is inferior to simple SVM and Random Forest models, with the available data. Hence, the approach is dropped for futher experiment.

#### **4.8 Current Machine Learning Model Architecture**

- App and Model version: 4.1
- Dataset Duration: From July 2022 to 5<sup>th</sup> Jan 2023
- Changes: Added new features in Dataset

- New Features: Count of Physical activity transition (running, walking, avg lock-unlock in 5 mins and 10 mins, Duration spent on various phone modes, travel as new app category)

Model Performance System:

Figure 4.13

Model Architecture (Emotion Prediction from User Data, Demographics, Personality)

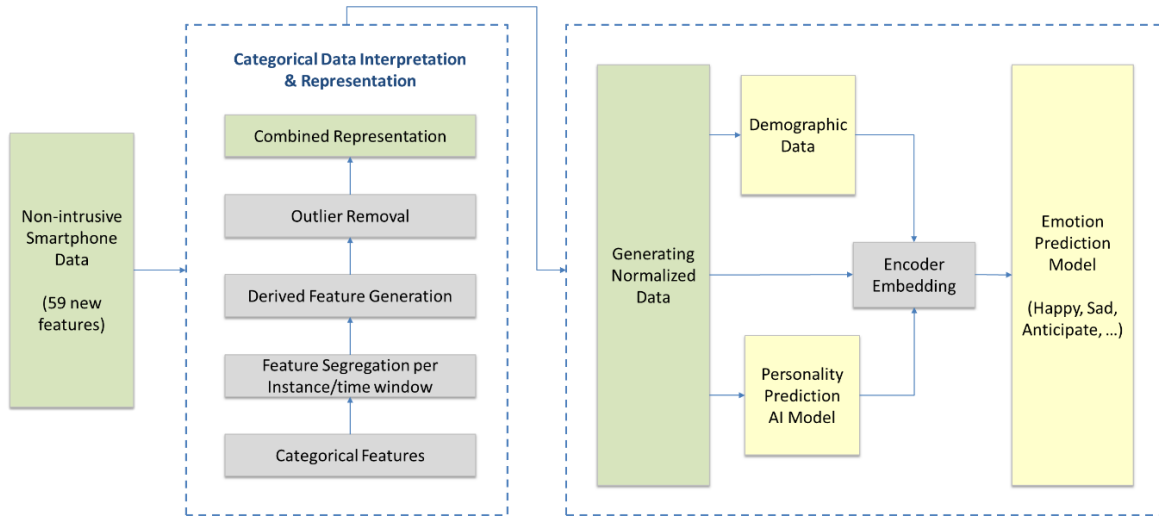


Table 4.8

Emotion Prediction: Without Demography and Personality

Model	Accuracy	F1-Score (Macro)
Random Forest	57	52
Catboost	66	56
SVM	70	53

Improvement in Version 4.1 over 4.0 due to new dataset:

*Table 4.9*

*Emotion Prediction Improvements: Version 4.0 and Version 4.1*

<b>Parameters</b>	<b>V4.0</b>	<b>V4.1</b>
Accuracy	51%	57.75%
F1-Score	46	51
Model Size (MB)	1.56	1.6
Inference Time (CPU)	350 ms	350 ms
End to End pipeline	420 ms	420 ms

#### **User Trial Performance for Version 4.0**

- Month – January
- Duration – 1 month

*Table 4.10*

*Emotion Model Results (Precision, Recall, F1-Score): Mar 2023*

<b>Emotion</b>	<b>Precision %</b>	<b>Recall %</b>	<b>F1-Score %</b>
Happy	80.35	95.07	87.09
Anticipate	93.75	57.69	71.42
Neutral	79.56	85.71	82.52
Sad	100	5	9.52
Tired/Sleepy	67.3	100	80.454
<b>Overall Accuracy %</b>	<b>79.57%</b>		

- Total accuracy: 79.57%
- Dataset duration: From July 2022 to 5<sup>th</sup> Oct 2022
- Unique users: 106
- Total dataset points: ~22K
- Labelled data points: ~2K (10%)
- System model version: 4.0 (includes demography and memory stats features)
- On-Device model version: 4.0 (excludes demography and memory stats features)
- Dataset duration: From July 2022 to 5<sup>th</sup> Oct 2022
- Dataset version: Version 1.0

### **Models:**

- Deployed model is Random Forest (RF) for both the task
- Number of classes: 5 emotions | 3 sentiments
- Emotions: Happy, Sad, Anticipate, Neutral, Tired/Sleepy
- Sentiments: Positive, Negative, Neutral

### **OnDevice Model KPIs:**

- Emotion accuracy: 51% | F1 score: 46% (RF Model)
- Sentiment Accuracy: 71% | F1 score: 66% (RF Model)
- Model Size: 1.59 MB (Emotion) | 535 KB (Sentiment)
- Preprocessing time: ~40 ms
- Emotion Inference Time: 350 ms (CPU)
- End to end pipeline (with personality): ~420 ms
- End to end pipeline (without personality): ~390 ms

- Not included: System memory, Demography

**System Model KPIs:**

- Emotion accuracy: 67.73% | F1 score: 56% | AP Score: 0.72 (RF Model)
- Sentiment Accuracy: 77.95% | F1 score: 78% | AP score: 0.82 (SVM Model)
- Model Size: ~3.8 MB (Emotion) | 494 KB (Sentiment)
- Includes Demography, System Memory Info

The detailed ablation study on Demography and personality features, various model comparisons, for both Emotion and Sentiment Prediction tasks are detailed and summarized in the next section (Chapter 5)

**4.9 Model Upgrades**

**4.9.1 Release stats – 16 May 2023 (Version 4.0, Model Version 1)**

*Table 4.11  
Emotion Model Results (Precision, Recall, F1-Score): May 2023*

<b>Emotion</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
Happy	84.98	93.34	88.96
Anticipate	84	56.75	67.74
Neutral	81.83	90.23	85.82
Sad	45.45	13.51	20.83
Tired/Sleepy	66.66	100	80
Other	0	0	0

Table 4.12  
Emotion Model - Confusion Matrix: May 2023

	Happy	Aniticipate	Neutral	Sad	Tired/Sleepy	Other
Happy	883	1	43	2	17	0
Anticipate	7	21	8	0	1	0
Neutral	45	0	536	3	10	0
Sad	17	0	14	5	1	0
Tired/Sleepy	0	0	0	0	86	0
Other	87	3	54	1	14	0

(Y-axis: Actual Values, X-Axis: Predicted Values)

**Latest stats on user feedback for emotion predictions (Version 4.0):**

- Total valid data points (neither predictions nor ground-truth have NA): 1859
- Correct Predictions: 1531
- Accuracy: 0.8235610543302851
- Total predictions (including those that do not have ground truth): 16185

**4.9.2 Release stats – 21 Jun 2023 (Version 4.1, Model Version 2)**

Table 4.13  
Emotion Model Results (Precision, Recall, F1-Score): Jun 2023

Emotion	Precision	Recall	F1-Score
Happy	85.13	84.81	84.97
Anticipate	80.00	66.67	72.72
Neutral	73.91	68.00	70.83



Sad	0	0	0
Tired/Sleepy	69.04	100	81.69
Other	0	0	0

*Table 4.14*  
*Emotion Model - Confusion Matrix: Jun 2023*

	Happy	Aniticipate	Neutral	Sad	Tired/Sleepy	Other
Happy	229	0	7	1	33	0
Anticipate	2	4	0	0	0	0
Neutral	6	0	51	0	18	0
Sad	7	0	2	0	4	0
Tired/Sleepy	0	0	0	0	174	0
Other	25	1	9	0	23	0

(Y-axis: Actual Values, X-Axis: Predicted Values)

**Latest stats on user feedback for emotion predictions (Version 4.1):**

- Total valid data points (neither predictions nor ground-truth have NA): 596
- Correct Predictions: 458
- Accuracy: 0.7684
- Total predictions (including those that do not have ground truth): 3803

The detailed ablation study, architecture comparison, results and analysis are captured in next chapter (Chapter 5).

CHAPTER V:  
EMOTION RECOGNITION: RESULTS AND ANALYSIS

### 5.1 Emotion Prediction: Result & Analysis

In this section, we discuss the results related to our annotation and emotion prediction methods. One significant point is that our method does not incur a higher mental workload than filling annotations using the widely-used Self-Assessment Manikin (SAM) method (Bradley et al., 1994).

We perform a detailed comparative study between prior works involving smartphone-based annotation techniques to perform various emotion prediction tasks as shown in Table 5.1. We observe that the Reflektions application fairs better in terms of preserving user privacy and capturing natural emotion as no private data is used and user smartphone-related tasks are not hampered. Additionally, we also make use of demographic and personality information given by the user to improve prediction results which is not explored by prior works. We are also able to collect sizeable annotated data samples without any monetary rewards for the participants.

*Table 5.1  
Comparative study between various smartphone-based annotation applications*

Features (O: Yes, X: No)	MoodExplorer	iSelf	MoodScope	EmoSensing	Reflektion (Proposed)
Collection methodology	Daily user tasks	Daily user tasks	Field study group	Recruited users	<b>Daily user tasks</b>
User feature (age, sex, etc.)	X	X	X	O	<b>O</b>
Use of personality information	X	X	X	X	<b>O</b>
Device Features (Bluetooth, Lock/Unlock, etc.)	O	O	O	X	<b>O</b>
User App Usage Behavior (Duration, category etc.)	O	O	O	O	<b>O</b>
Private Content (SMS, Location, Images, etc.)	O	O	O	O	<b>X (Private content not used)</b>
Monetary Rewards for participants	O	X	O	O	<b>X</b>
Number of Participants	30	10	32	27	<b>100</b>
Number of Data samples	X	3600	X	X	<b>13700</b>
Number of Annotated samples	X	3600	X	X	<b>1532</b>
Task Type	Compound Emotion	Emotion	Mood	Multiclass Emotion	<b>Emotion, Sentiment</b>

Since there is no public dataset available for benchmarking emotion and sentiment prediction using mobile phone usage data, we could not provide a comparison of our method performance with respect to State of the art (SOTA) and instead validated our dataset using standard models. For our labelled dataset, we train both machine learning (ML) and deep learning (DL) models to classify emotions and sentiments. Since our labelled dataset is small, ML models are found to be more effective than DL models. As shown in Table 5.2, we train several tree-based models such as Random Forest (Breiman, 2001), Support Vector Machine (SVM) – Radial Basic Function (RBF) Kernel (Hearst et al., 1998), XGBoostClassifier (Chen et al., 2016), Gradient Boosting Classifier (Chen et al., 2016), CatBoost (Dorogush et al., 2018) and Light Gradient Boosting Machine (LightGBM) (Ke et al., 2017) for our tasks. We divide the dataset into 80:20 training and testing ratio, with 20% of the data points used for testing. For every trained model, we apply grid search-based hyperparameter tuning and observe the performance improvement in nearly all the models. For fine tuning tree-based models, mostly two parameters are considered: max depth and n estimators which represent the depth of the tree and the number of trees respectively. As shown in Table 5.2, we get the best performance for the emotion prediction task using Random Forest with the following hyperparameters: max depth=18; n estimators=300; min samples leaf =5. Similarly, for the sentiment task, SVM with hyperparameters as gamma=0.009; kernel = 'rbf'; C =3 provide the best performance.

For comparison, we also train two deep learning models namely Multi-Layer Perceptron (MLP) and 1D convolution. MLP consists of three dense layers with units viz 32, 64, and 128 in sequence. Similarly, the 1D convolution model consists of three 1DConv layers having kernel sizes of 32, 64, and 128 followed by a 128-dense layer. The last layer for both the models is the classification layer with 5 units for the emotion task and 3 units

for the sentiment task. For both models, we use the activation function as Rectified Linear Unit (ReLU) for intermediate layers and SoftMax for the classification layer. For hyperparameter tuning and to get an optimum number of layers for both models, we use the Keras Tuner framework, and above discussed configuration provided the best performance. Other hyperparameters are as follows: learning rate = .0001; batch size = 64; epochs = 100.

*Table 5.2  
Performance of different classifiers on Emotion and Sentiment prediction tasks*

Model	Emotion Prediction			Sentiment Prediction		
	Accuracy (%)	F1 Score (%) (Macro Average)	AP Score (Micro)	Accuracy (%)	F1 Score (%) (Macro Average)	AP Score (Micro)
SVM (RBF)	65.21	53	.73	<b>77.95</b>	<b>78</b>	.82
XGBClassifier	51.71	43	.48	72.60	71	.81
Gradient Boosting Classifier	49.42	39	.46	73.05	73	.80
CatBoost	66.59	55	.70	73.49	72	.83
LGBClassifier	52.40	45	.51	73.49	73	.82
Random Forest	<b>67.73</b>	<b>56</b>	.72	71.26	71	.80
MLP (3 Layer)	62.92	52	.64	73.49	74	.82
1D Convolution (4 Layer)	61.09	48	.70	71.26	67	<b>.84</b>

Since our training and validation data is highly imbalanced (Happy: 45%, Neutral: 33%, Anticipate: 8%, Sad: 5%, Tired/Sleepy: 9%), we calculate F1 score, Average Precision (AP) for both the sentiment (3 classes: positive, negative, and neutral) and emotion (5 classes: Anticipate, Happy, Neutral, Sad and Tired/Sleepy) prediction tasks. F1 score (macro average) is a simple average over all classes, so each class is given equal weight independent of their proportion and AP (Micro) takes into account the class imbalance in calculating average precision. Table 5.2 shows the performance of several models for both tasks. For the emotion prediction task, the Random Forest algorithm provides the best performance among others achieving 67.73% accuracy, 56 % F1 -score,

and an AP score of .72. The SVM (RBF kernel) outperformed other models with an accuracy of 77.95%, F1-score of 78% on the sentiment prediction task.

*Table 5.3  
Performance results for Emotion and Sentiment prediction tasks*

Metrics	Emotion Prediction (Random Forest)					Sentiment Prediction (SVM)		
	Anticipate	Happy	Neutral	Sad	Tired/Sleepy	Negative	Neutral	Positive
Precision	49	70	76	55	53	80	70	83
Recall	59	87	56	26	46	85	72	79
F1 Score	53	78	64	35	49	82	71	81

Table 5.3 provides class-wise precision, recall, and F1-score by best performing Random Forest and SVM models on emotion and sentiment prediction tasks respectively. Low scores for Sad and Tired/Sleepy emotions can be attributed to the lower number of samples in the data distribution.

*Table 5.4  
Ablation Study for Emotion Task*

Model	Demography	Personality	F1 Score (%) (Macro Average)	AP Score (Micro)
SVM (RBF)	X	X	51	.72
	O	X	53	.73
	<b>X</b>	<b>O</b>	<b>54</b>	<b>.74</b>
	O	O	53	.73
Random Forest	X	X	48	.64
	O	X	49	.65
	X	O	55	.71
	<b>O</b>	<b>O</b>	<b>56</b>	<b>.72</b>

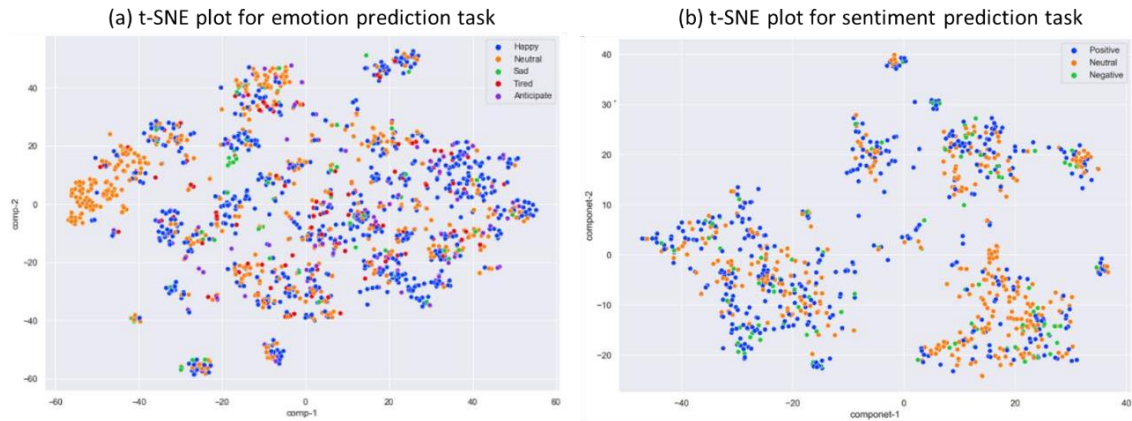
We perform an ablation study to analyse and understand the impact of demography (age, gender) and personality (five personality traits) as input features to the emotion model. Table 5.4 discuss experimentation performed for top-performing ML models for

the emotion prediction task. It can be seen that there is always performance improvement whenever either demography or personality or both are taken as input to the model. Another interesting observation is, when personality alone is considered as input (dropping demography features) along with other features, we get the best performance for SVM and fair performance on the Random Forest model. It establishes the fact that there is implicit relation between emotion and personality, and knowing the user's personality can be beneficial in determining the user's emotional state. We also observed a little performance drop when both demography and personality features are considered. One possible reason can be high data imbalance with respect to gender and age group (refer Figure 4.5) leading to downward performance.

We notice subpar performance for the emotion prediction task in comparison to the sentiment prediction task. It is particularly due to 1) the relatively lesser number of samples for the classes such as Anticipate, Sad, and Tired/Sleepy. 2) feature set for rare classes like Sad, Tired/Sleepy, etc., is highly overlapping with classes such as happy and neutral as demonstrated in the t-SNE plot in Figure 5.1a and thus making it challenging for the models to learn exact boundaries. Relatively better performance for the sentiment prediction task can be attributed to the reduced complexity of the task as a result of grouping emotions into sentiments as confirmed by the t-SNE plot in Figure 5.1b. The t-SNE plot shows that both positive and neutral class data points form separate clusters (except for a few outliers) while negative class data points lie mostly in the periphery of the other two classes making the model learn the pattern and perform better.

Figure 5.1

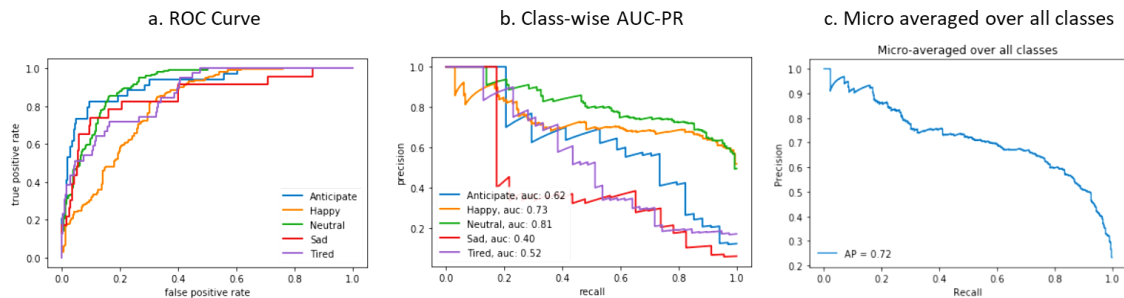
*t-SNE plot for emotion and sentiment prediction tasks using categorical features*



In Figure 5.2, we demonstrate various performance metrics for Random Forest on emotion prediction task. Figure 5.2a shows the True Positive Rate (TPR) versus False Positive Rate (FPR) graph, called as Area Under the Curve-Receiver Operating Characteristic (AUC-ROC) curve, across varying thresholds for target emotion classes. The ROC curve with better discrimination ability lies top left corner. It can be seen that Neutral and Anticipate have greater discriminate abilities than Sad. Figure 5.2b shows class-wise Precision-Recall (PR) performance (referred as class-wise AUC-PR curve). For each class, it is calculated by considering the problem statement as binary and represents the discrimination capacity of individual class versus others. In our case, we perform target class vs others analysis to get class wise average precision score. It can be seen that Happy and Neutral perform comparatively better and Sad has the lowest precision score. Figure 5.2c shows a trade-off between precision and recall for our emotion prediction task, micro-averaged over all the classes.

Figure 5.2

Performance of Random Forest classifier on Emotion Prediction task



Given the challenges of designing data collection applications for smartphones, there are a few limitations to our work. First, the Reflektion app is designed to capture real-world smartphone interactions non-intrusively and strictly adhere to user privacy. This prohibits us to use user content like chat messages, pictures etc. which if utilized can naturally increase the accuracy. Still, we believe our findings provide a first step towards collecting more precise emotional ground-truth labels without compromising user privacy. Second, while we designed and iterated alternatives for inputting real-time emotion annotation, we explored external factors/activities influencing emotion in a limited way. However, our aim here was to firstly validate how well our designed annotation method works in comparison to the standard practice of asking participants to log emotions specifically. Further, we also observed imbalance in our dataset (Gender and age-group, emotion data distribution), which negatively impacted model predictions. However, we overcome these limitations during the preprocessing phase by using appropriate feature engineering strategies.



## CHAPTER VI:

### DISCUSSION

#### 6.1 Research Question One

*What are the State-of-the-art model architectures, that are feasible for commercialization and real-time applications, and that can be inferred with low computation devices (like smart phones)?*

- Chapter II elaborates about existing On-Device state-of-the-art model architectures and available data collection methodologies.
- Section 2.2 explains about different modalities (text, audio, vision, video, sensor and multimodal data), using which emotion is detected/recognized. This is summarized in Table 2.1. It is clear from the prior arts that emotion understanding from smartphone user activities and demographics are not available and there exists a need for coming up with novel AI/ML architectures to detect emotion from the same.
- Section 2.3 explains various data collection methods, summarized in Table 2.2. Most of these prior works rely on the content of user activity, like message content which may contain user-private information. Further, most studies collect data by incentivizing the user to record emotions which may result in unnatural and forced data. These drawbacks need to be addressed in our work and they establish a clear need to propose non-intrusive smartphone activity annotation technique followed by some significant insights.

#### 6.2 Research Question Two

*How can the data be collected - demographic data, user profile, user's device state, user activities in the devices?*

- **Objective - Emotion Data collection Experience:** Prepare methods for data collection, and design emotion data collection experience.
- This is explained in Section 4.1. The user trial application called Reflektion is proposed, designed and developed to collect user data from beta users. This is mainly to understand user's pattern related to emotions, to analyse and bring emotion based insights from the users, and to build Emotion AI models based on Non-intrusive data.
- The application process includes developing intuitive beta application (as shown in Figure 4.1), methods to get user consents (as shown in Appendix B), periodic reminders to enter the emotion details beyond direct entry in the application (as shown in Figure 4.2), and finally tagging the user data with user logged emotion and uploading it to internal server.
- Appendix B summarizes the survey that is taken while onboarding the users in Reflektions user trial application, to understand user's Demographics and Personality traits. Appendix C depicts about the application permissions dialog and the informed consent received from the user, while onboarding the user trial application.

### 6.3 Research Question Three

*How can we mitigate privacy issues while collecting above mentioned data from the user and while processing user data for emotion recognition tasks?*

- **Objective - System Design and Data collection:** Propose a novel annotation technique using the smartphone for generating ground truth labels. Develop a user trial app based on the same and distribute it to the participants for collecting tagged ground truth data. Identify personally identifiable information (PII) to protect privacy and incorporate it while collecting data.

- Section 4.2 explains about design and implementation of an effective system functionality. The application system design has mainly three core components shown in the Figure 4.3: 1) Work Manager, 2) Monitors, and 3) Logger.
- In particular, the monitor involves capturing right emotion specific user data. It includes various broadcast receivers and listeners that help capture significant user activities like App usage time, Wi-Fi connection/disconnection, Flight Mode On/Off, etc. Some features like Call duration are further used to derive essential elements like average call duration for the top 10 most contacted individuals. These are non-intrusive features collected through user consent. We ensure to mask any Personally Identifiable Information (PII) if present in any collected user data.
- **Important points:**
  - No PII is collected.
  - No 3<sup>rd</sup> party app data is parsed or extracted.
  - SMS contents are not read.
  - Adequate measure taken so that data cannot be traced back to originator (Encryption, Anonymization, Data –Deid)
  - The data collected split into train and test set and used to build a machine learning model. The data will not be used for marketing or any other purpose.
- **Following are the direct data collected, where privacy is ensured:**
  - Bluetooth(On/Off, Pairing/unpairing, Device discovery, Device connected/disconnected)
  - Contacts( Missed call, incoming/outgoing call: count, duration)
  - App state change ( install/ uninstall/update/replaced)

- Lockscreen (lock/unlock)
  - Wifi (Scan results, on/off, network connected/disconnected)
  - SMS (Status)
  - Physical activity (vehicle in, walking, running, on bicycle, still)
  - Phone Mode (Flight Mode, Silent, Vibrate, Normal)
  - Music (Metadata changed)
  - App Usage (Top 10 app packages used, duration) : Once per window
  - Calendar (Event info , event time) : Once per window
  - System Parameters ( Battery percentage, RAM, ROM)
- Finally, a light-weight AI/ML model is built to predict emotion in user’s device so that the processing and inference happens on-device with the user data and the user data does not go out of the user’s smartphone.
  - To summarize, only the non-private and non-intrusive user data is collected for analyzing user emotions, with the necessary user consents. Any emotion understanding with the user data happens only on-device and the user data does not leave the user’s smartphone, to ensure user privacy.
  - Further, Appendix D shows privacy policies that are added as part of Terms and Conditions while onboarding the user in user trial application.

#### 6.4 Research Question Four

*What are the insights from the collected data, before performing emotion recognition task?*

- **Objective - Emotion Data analysis:** Derive critical insights from the data collected from the user trial app about emotion, personality, and user behavior.

- This is elaborated in section 4.3. The section outlines the Data Processing Pipeline (as shown in Figure 4.4).
- At first, the section explains various data distribution:
  - Ground Truth Labels Tagging & Demographic Data Distribution (Gender and age group distribution), as shown in Figure 4.5.
  - Distribution of emotions collected from the users, as shown in Figure 4.6
  - Distribution of app category launch and duration, as shown in Figure 4.7
  - Distribution of emotions in different category of applications (App Usage vs Emotion), as shown in Figure 4.8. For example, Emotion in music apps, entertainment apps, social category apps and so on.
- And further, various emotion-driven insights are identified and presented, as shown below:
  - Emotion Distribution among various smart phone features, as shown in Figure 4.9. Example: When user is in 'Anticipation' emotion, Users tend to lock/unlock phones more and food/drink category app usage is high. This can be when user orders food, keeps lock/unlock the phone and enters into anticipation state. Similar emotion driven insights are explained in this section.
  - Further, Chart representing the Sequence of Emotion entries for few Participants are also researched and discussed, as shown in Figure 4.10. Example: user tend to be in 'Surprise' emotion soon after 'Anticipate' emotion. Such insights related to sequence of emotions are presented.

## **6.5 Research Question Five**

*What should be the model architecture for categorical data collected for emotion recognition and how should the end-to-end pipeline look like? (Data -> Novel Fusion*

*Embedding -> unique fusion-model-architecture -> KPI evaluation for real-time applications (Inference) -> Beta Deployment).*

- **Objective - Feature Engineering & Emotion Prediction:** Design a model having the relation between personality, emotion, and user smartphone behavior in a non-intrusive way, maintaining user privacy.
- The overview of Model Prediction Pipeline, utilizing Novel smartphone data is presented in Section 4.4. The overview of this pipeline is shown in Figure 4.11.
- In this section, various adopted feature engineering techniques are described, such as Feature selection, Removing outliers, Normalizing the features and finally encoding the normalized features. Further, techniques for handling class-imbalance for user activity data are discussed.
- The classification models such as Random Forest Classifier and SVM Classifier, along with the results are detailed. These are mainly used to predict emotions such as Happy, Neutral, Anticipate, Sad, Tired/Sleepy. Apart from emotion prediction, we also provide predictions for three sentiments polarity viz. positive, negative, and neutral on the same dataset. The two deep learning models are also trained, namely 3-layer Multi-Layer Perceptron (MLP) model and 4-layer 1D convolution, for emotion and sentiment prediction tasks.
- We observe that the Reflektions application fairs better in terms of preserving user privacy and capturing natural emotion as no private data is used and user smartphone-related tasks are not hampered. Additionally, we also make use of demographic and personality information given by the user to improve prediction results which is not explored by prior works. We are also able to collect sizeable annotated data samples without any monetary rewards for the participants.

## 6.6 Research Question Six

*What is the current state-of-the-art (SOTA) KPI and what are the optimal KPIs that should be considered for the model?*

- **Objective - Result and Analysis:** Validate the collected non-intrusive data by developing and implementing a system for automatic emotion detection, extract different features from the collected tagged data, train the machine learning model, test its performance and compare with SOTA KPIs.
- The Performance of different classifiers on Emotion and Sentiment prediction tasks are developed, analyzed and benchmarked. The list of classifier models include SVM (RBF kernel), XGB Classifier, Gradient Boosting Classifier, CatBoost, LGB Classifier, Random Forest, MLP (3 Layers), 1D Convolution network (4 layers). Random Forest and SVM Classifier outperforms other classification models.
- For the emotion prediction task, the Random Forest algorithm provides the best performance among others achieving 67.73% accuracy, 56 % F1 -score, and an AP score of .72. The SVM (RBF kernel) outperformed other models with an accuracy of 77.95%, F1-score of 78% on the sentiment prediction task, as shown in Table 5.2. Further, Table 5.3 provides class-wise precision, recall, and F1-score by best performing Random Forest and SVM models on emotion and sentiment prediction tasks respectively.
- We perform an ablation study to analyse and understand the impact of demography (age, gender) and personality (five personality traits) as input features to the emotion model. Table 5.4 discuss experimentation performed for top-performing ML models for the emotion prediction task. It can be seen that there is always performance improvement whenever either demography or personality or both are taken as input to the model.

- Another interesting observation is, when personality alone is considered as input (dropping demography features) along with other features, we get the best performance for SVM and fair performance on the Random Forest model. It establishes the fact that there is implicit relation between emotion and personality, and knowing the user's personality can be beneficial in determining the user's emotional state.

### **6.7 Summary of Findings**

- First work to capture the relation between personality, emotion, and user smartphone behavior in a non-intrusive way, maintaining user privacy.
- Proposed novel annotation technique using the smartphone for generating ground truth labels. Developed user trial app based on the same and distributed it to over 100 participants for collecting tagged ground truth data | Optional and Automated
- Derived critical insights with the data collected from the user trial app, about emotion, personality, and user behavior.
- Validated the non-intrusive data by developing emotion prediction system. Extracted different features from the collected tagged data, trained the ML model and tested its performance. Achieved SOTA accuracy of 67.73% for Emotion and 77.95% for Sentiment, respectively.



## CHAPTER VII: SUMMARY AND RECOMMENDATIONS

### **7.1 Conclusion**

A design for a real-time emotion annotation technique for smartphone usage behavior without invading user privacy, is presented. The proposed approach enables researchers to collect emotion annotations while using a smartphone. Through collection methodology, our method ensures that it does not incur extra pressure to log emotions, as emotion entry is optional. This also ensures that the logged data capture genuine emotions. Moreover, the consistency of annotations is verified by deriving new insights about user behavior and strengthening insights derived from previous work in the field. Further the annotated data is used to predict user emotions and sentiment with good accuracy. It is also demonstrated that using the user personality as one of the features, emotion prediction accuracy can be enhanced. Once user emotion is known, one can enable various novel on-device and multi-device experiences. It is possible to achieve deeper personalization in applications like Music players, Search, Contacts, etc. on the device where recommendations can take user emotion into account. The proposed research work underscores the importance of collecting ground truth emotion annotations, which is essential for ensuring accurate user emotion recognition non-intrusively.

### **7.2 Recommendations for Future Research**

In the future, the plan is to explore deep learning methodologies like zero-shot learning to increase the accuracy of emotion prediction. Also, we intend to distribute a mobile application across geographies and a few educational institutions to improve the demographic distribution of data and to achieve better generalization.

Further, the future research could include coming up with SOTA DNN architecture, for emotion, personality and sentiment prediction tasks, utilizing non-private categorical data. This involves novel model architecture for categorical data structure. Further, building the Knowledge Graph can help in analyzing patterns and building relationships among the smartphone features. This can bring advanced and very useful user-needed features, such as providing knowledge based personalized recommendations. Recently, with the evolving Generative AI techniques, such as Large Language Model (LLM) and Large Multimodal Model (LMM), it is possible to provide meaningful and most advanced recommendations and insights to the users.

APPENDIX A

IBIS 2023 CONFERENCE

This work is accepted and published as a Research Paper in International Conference on Business and Integral Security 2023 (IBIS 2023).

*Figure A.1*  
*IBIS 2023 Conference: Presentation Certificate*



Paper is available at: <https://www.gbis.ch/index.php/gbis/article/view/261>

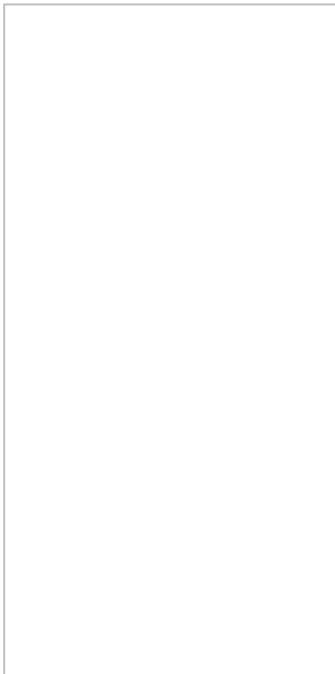
Authors: Barath Raj Kandur Raja, Sumit Kumar, Prof Mario Silic et al.

## APPENDIX B

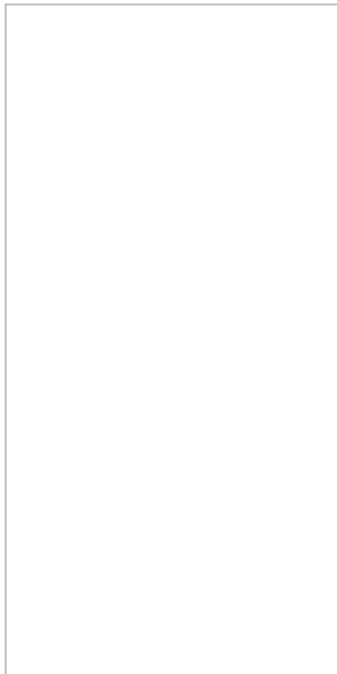
### SURVEY: DEMOGRAPHICS & PERSONALITY TRAITS

The below survey is taken while onboarding the users in Reflektions user trial application, to understand user's Demographics and Personality traits.

*Table B.1*  
*Survey: Demographics*



*Table B.2*  
*Survey: Demographics Filled*



*Table B.3*  
*Survey: Personality*

A screenshot of a mobile application survey titled "Reflektions". The header says "Help us know you better". The survey consists of five personality trait questions, each with a description and three radio button options: Low, Medium, and High. The traits are: Openness (Degree of intellectual curiosity and preference for novelty), Agreeableness (Measure of one's trusting and helpful nature), Extraversion (Tendency to seek company and talk to others), Conscientiousness (Tendency to be organized), and Neuroticism (Ability to handle stress). At the bottom, there are two buttons: "Skip" (light gray) and "Next" (red).

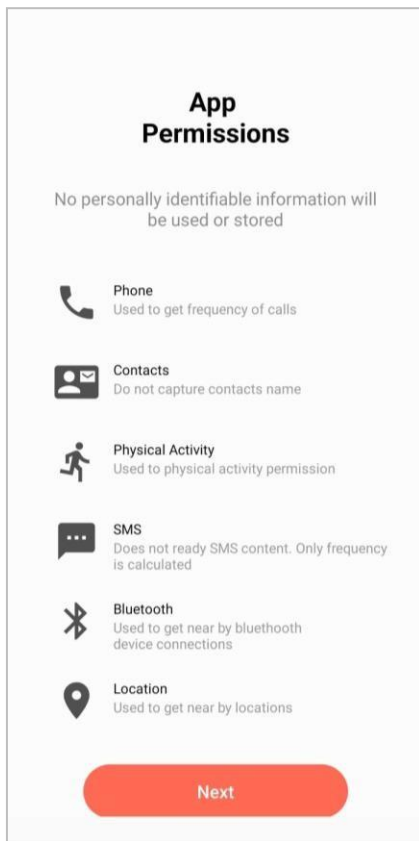
## APPENDIX C

### INFORMED CONSENT: PERMISSIONS

The below permissions dialog is shown and informed consent from the user is received, while onboarding the user trial application.

*Table C.1*

*Informed Consent: App Permissions (No Personally Identifiable Information)*

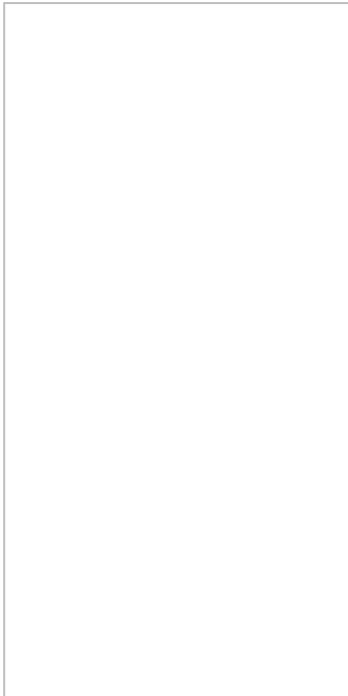


## APPENDIX D

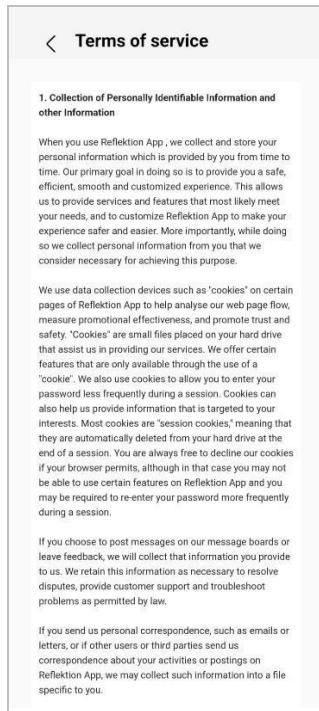
### INFORMED CONSENT: PRIVACY POLICY

The below privacy policies are added as part of Terms and Conditions while onboarding the user in user trial application.

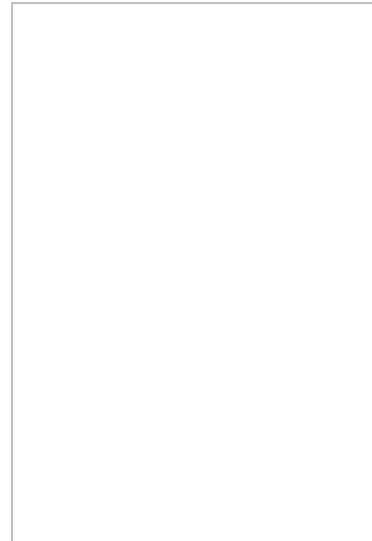
*Table D.1  
Informed Consent:  
Onboarding screen*



*Table D.2  
Informed Consent: Privacy  
Policy – Screen 1*



*Table D.3  
Informed Consent: Privacy  
Policy – Screen 2*



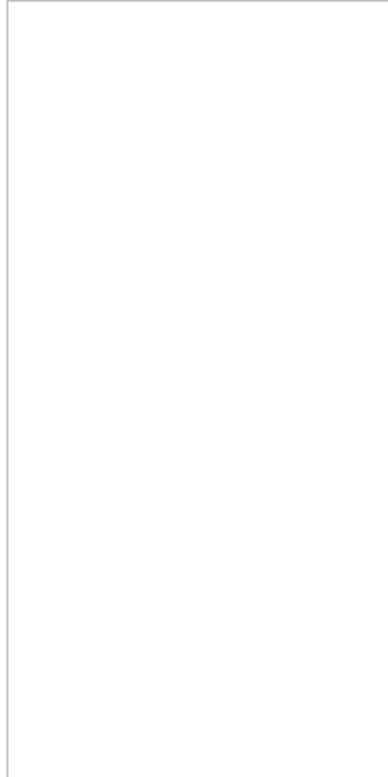
*Table D.4*

*Informed Consent: Privacy Policy – Screen 3*



*Table D.5*

*Informed Consent: Privacy Policy - Screen 4*



## REFERENCES

- PS, S. and Mahalakshmi, G., (2017). "Emotion models: a review". *International Journal of Control Theory and Applications*, 10(8), pp.651-657.
- Felbo, B., Mislove, A., Søggaard, A., Rahwan, I. and Lehmann, S., (2017). "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm". *arXiv preprint arXiv:1708.00524*.
- Shen, W., Wu, S., Yang, Y. and Quan, X., (2021). "Directed acyclic graph network for conversational emotion recognition". *arXiv preprint arXiv:2105.12907*.
- Majumder, N., Poria, S., Hazarika, D., Mihalcea, R., Gelbukh, A. and Cambria, E., (2019, July). "Dialoguernn: An attentive rnn for emotion detection in conversations". In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 6818-6825).
- Ghosal, D., Majumder, N., Poria, S., Chhaya, N. and Gelbukh, A., (2019). "Dialoguecn: A graph convolutional neural network for emotion recognition in conversation". *arXiv preprint arXiv:1908.11540*.
- Issa, D., Demirci, M.F. and Yazici, A., (2020). "Speech emotion recognition with deep convolutional neural networks". *Biomedical Signal Processing and Control*, 59, p.101894.
- Krishnan, P.T., Joseph Raj, A.N. and Rajangam, V., (2021). "Emotion classification from speech signal based on empirical mode decomposition and non-linear features: Speech emotion recognition". *Complex & Intelligent Systems*, 7, pp.1919-1934.
- Minaee, S., Minaei, M. and Abdolrashidi, A., (2021). "Deep-emotion: Facial expression recognition using attentional convolutional network". *Sensors*, 21(9), p.3046.
- Yue-Hei Ng, J., Hausknecht, M., Vijayanarasimhan, S., Vinyals, O., Monga, R. and Toderici, G., (2015). "Beyond short snippets: Deep networks for video classification". In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4694-4702).



- Carreira, J. and Zisserman, A., (2017). “Quo vadis, action recognition? a new model and the kinetics dataset”. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 6299-6308).
- Xie, S., Sun, C., Huang, J., Tu, Z. and Murphy, K., (2018). “Rethinking spatiotemporal feature learning: Speed-accuracy trade-offs in video classification”. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 305-321).
- Akbari, H., Yuan, L., Qian, R., Chuang, W.H., Chang, S.F., Cui, Y. and Gong, B., (2021). “Vatt: Transformers for multimodal self-supervised learning from raw video, audio and text”. *Advances in Neural Information Processing Systems*, 34, pp.24206-24221.
- Brave, S. and Nass, C., (2007). “Emotion in human-computer interaction”. In *The human-computer interaction handbook* (pp. 103-118). CRC Press.
- Ooi, C.S., Seng, K.P., Ang, L.M. and Chew, L.W., (2014). “A new approach of audio emotion recognition”. *Expert systems with applications*, 41(13), pp.5858-5869.
- Akhand, M.A.H., Roy, S., Siddique, N., Kamal, M.A.S. and Shimamura, T., (2021). “Facial emotion recognition using transfer learning in the deep CNN”. *Electronics*, 10(9), p.1036.
- You, Q., Luo, J., Jin, H. and Yang, J., (2016, February). “Building a large scale dataset for image emotion recognition: The fine print and the benchmark”. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 30, No. 1).
- Kanjo, E., Younis, E.M. and Ang, C.S., (2019). “Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection”. *Information Fusion*, 49, pp.46-56.
- Majumder, N., Poria, S., Hazarika, D., Mihalcea, R., Gelbukh, A. and Cambria, E., (2019, July). “Dialoguernn: An attentive rnn for emotion detection in conversations”. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 6818-6825).
- Revelle, W. and Scherer, K.R., (2009). “Personality and emotion”. *Oxford companion to emotion and the affective sciences*, 1, pp.304-306.

- Busso, C., Bulut, M., Lee, C.C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J.N., Lee, S. and Narayanan, S.S., (2008). "IEMOCAP: Interactive emotional dyadic motion capture database". *Language resources and evaluation*, 42, pp.335-359.
- Livingstone, S.R. and Russo, F.A., (2018). "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*, 13(5), p.e0196391.
- R Fulcher. (2022). *Create intuitive and beautiful products with Material Design*. Available at: <https://material.io/design>
- Chloe Blanchard. (2022). *Mobile App Design Fundamentals: 10 tips for an effective content strategy*. Available at: <https://clearbridgemoible.com/mobile-app-design-fundamentals-10-tips-for-an-effective-content-strategy/>.
- Hoefel F Francis T. (2022). *Generation Z and its implications for companies*. Available at: <https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/true-gen-generation-z-and-its-implications-for-companies>.
- Yus, F., (2014). "Not all emoticons are created equal". *Linguagem em (Dis) curso*, 14, pp.511-529.
- Morshed, M.B., Saha, K., Li, R., D'Mello, S.K., De Choudhury, M., Abowd, G.D. and Plötz, T., (2019). "Prediction of mood instability with passive sensing." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3), pp.1-21.
- Bogomolov, A., Lepri, B., Ferron, M., Pianesi, F. and Pentland, A., (2014, November). "Daily stress recognition from mobile phone data, weather conditions and individual traits." In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 477-486).
- Caldeira, C., Chen, Y., Chan, L., Pham, V., Chen, Y. and Zheng, K., (2017). "Mobile apps for mood tracking: an analysis of features and user reviews." In *AMIA Annual Symposium Proceedings* (Vol. 2017, p. 495). American Medical Informatics Association.
- Darvariu, V.A., Convertino, L., Mehrotra, A. and Musolesi, M., (2020). "Quantifying the relationships between everyday objects and emotional states through deep learning

- based image analysis using smartphones.” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1), pp.1-21.
- data.ai (2022). *The State of Mobile in 2022: How to Succeed in a Mobile-First World As Consumers Spend 3.8 Trillion Hours on Mobile Devices*. Available at: <https://www.data.ai/en/insights/market data/state-of-mobile-2022/>.
- Dixon, T., (2012). “Emotion: The history of a keyword in crisis.” *Emotion Review*, 4(4), pp.338-344.
- Donovan, R., Johnson, A., deRoiste, A. and O'Reilly, R., (2021). “A Multimodal Workflow for Modeling Personality and Emotions to Enable User Profiling and Personalisation.” In *VISIGRAPP (2: HUCAPP)* (pp. 145-152).
- Ekman, P., (1992). “An argument for basic emotions.” *Cognition & emotion*, 6(3-4), pp.169-200.
- Ferdinando, H. and Alasaarela, E., (2018, January). “Enhancement of emotion recognition using feature fusion and the neighborhood components analysis.” In *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2018)*. SCITEPRESS Science And Technology Publications.
- Furner, C.P., Racherla, P. and Babb, J.S., (2014). “Mobile app stickiness (MASS) and mobile interactivity: a conceptual model.” *The Marketing Review*, 14(2), pp.163-188.
- Hamari, J., Koivisto, J. and Sarsa, H., (2014, January). “Does gamification work?--a literature review of empirical studies on gamification.” In *2014 47th Hawaii international conference on system sciences* (pp. 3025-3034). Ieee.
- Hung, G.C.L., Yang, P.C., Chang, C.C., Chiang, J.H. and Chen, Y.Y., (2016). “Predicting negative emotions based on mobile phone usage patterns: an exploratory study.” *JMIR research protocols*, 5(3), p.e5551.
- John, O.P. and Srivastava, S., (1999). “The Big-Five trait taxonomy: History, measurement, and theoretical perspectives.”
- Lavid Ben Lulu, D. and Kuflik, T., (2013, March). “Functionality-based clustering using short textual description: Helping users to find apps installed on their mobile device.”

- In *Proceedings of the 2013 international conference on Intelligent user interfaces* (pp. 297-306).
- LiKamWa, R., Liu, Y., Lane, N.D. and Zhong, L., (2013, June). "Moodscope: Building a mood sensor from smartphone usage patterns." In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services* (pp. 389-402).
- McAdams, D.P., (2015). *The art and science of personality development*. Guilford Publications.
- McCrae, R.R. and Costa, P.T., (2008). "Empirical and theoretical status of the five-factor model of personality traits." *The SAGE handbook of personality theory and assessment, 1*, pp.273-294.
- Pelteret, M. and Ophoff, J., (2016). "A review of information privacy and its importance to consumers and organizations." *Informing Science, 19*, pp.277-301.
- Plutchik, R., (2001). "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice." *American scientist, 89*(4), pp.344-350.
- Roshanaei, M., Han, R. and Mishra, S., (2017, July). "Emotionsensing: Predicting mobile user emotion." In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017* (pp. 325-330).
- Russell, J.A. and Mehrabian, A., (1977). "Evidence for a three -factor theory of emotions." *Journal of research in Personality, 11*(3), pp.273-294.
- Bakker, I., Van Der Voordt, T., Vink, P. and De Boon, J., (2014). "Pleasure, arousal, dominance: Mehrabian and Russell revisited." *Current Psychology, 33*, pp.405-421.
- Sadeghian, A. and Kaedi, M., (2021). "Happiness recognition from smartphone usage data considering users' estimated personality traits." *Pervasive and Mobile Computing, 73*, p.101389.
- Schobel, J., Schickler, M., Pryss, R., Maier, F. and Reichert, M., (2014). "Towards process-driven mobile data collection applications: Requirements, challenges, lessons learned."

- Sun, B., Ma, Q., Zhang, S., Liu, K. and Liu, Y., (2017). "iSelf: Towards cold-start emotion labeling using transfer learning with smartphones." *ACM Transactions on Sensor Networks (TOSN)*, 13(4), pp.1-22.
- Zhang, X., Li, W., Chen, X. and Lu, S., (2018). "Moodexplorer: Towards compound emotion detection via smartphone sensing." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4), pp.1-30.
- Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P., (2002). "SMOTE: synthetic minority over-sampling technique." *Journal of artificial intelligence research*, 16, pp.321-357.
- Gonçalves, V.P., Giancristofaro, G.T., Filho, G.P., Johnson, T., Carvalho, V., Pessin, G., Neris, V.P.D.A. and Ueyama, J., (2017). "Assessing users' emotion at interaction time: a multimodal approach with multiple sensors." *Soft Computing*, 21, pp.5309-5323.
- Bradley, M.M. and Lang, P.J., (1994). "Measuring emotion: the self-assessment manikin and the semantic differential." *Journal of behavior therapy and experimental psychiatry*, 25(1), pp.49-59.
- Breiman, L., (2001). "Random forests." *Machine learning*, 45, pp.5-32.
- Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J. and Scholkopf, B., (1998). "Support vector machines." *IEEE Intelligent Systems and their applications*, 13(4), pp.18-28.
- Chen, T. and Guestrin, C., (2016, August). "Xgboost: A scalable tree boosting system." In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- Dorogush, A.V., Ershov, V. and Gulin, A., (2018). "CatBoost: gradient boosting with categorical features support." *arXiv preprint arXiv:1810.11363*.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.Y., (2017). "Lightgbm: A highly efficient gradient boosting decision tree." *Advances in neural information processing systems*, 30.
- Shapsough, S., Hesham, A., Elkhazraty, Y., Zualkernan, I.A. and Aloul, F., (2016, September). "Emotion recognition using mobile phones." In *2016 IEEE 18th*

- International Conference on e-Health Networking, Applications and Services (Healthcom)* (pp. 1-6). *IEEE*.
- Sarsenbayeva, Z., Marini, G., van Berkel, N., Luo, C., Jiang, W., Yang, K., Wadley, G., Dingler, T., Kostakos, V. and Goncalves, J., (2020, April). “Does smartphone use drive our emotions or vice versa? A causal analysis.” In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-15).
- Chen, W., Feng, F., Wang, Q., He, X., Song, C., Ling, G. and Zhang, Y., (2021). “Catgcn: Graph convolutional networks with categorical node features.” *IEEE Transactions on Knowledge and Data Engineering*.
- Wang, Z., Zhang, W., Liu, N. and Wang, J., (2021). “Scalable rule-based representation learning for interpretable classification.” *Advances in Neural Information Processing Systems*, 34, pp.30479-30491.
- Thost, V. and Chen, J., (2021). “Directed acyclic graph neural networks.” *arXiv preprint arXiv:2101.07965*.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V., (2019). “Roberta: A robustly optimized bert pretraining approach.” *arXiv preprint arXiv:1907.11692*.
- Tsai, Y.H.H., Bai, S., Liang, P.P., Kolter, J.Z., Morency, L.P. and Salakhutdinov, R., (2019, July). “Multimodal transformer for unaligned multimodal language sequences.” In *Proceedings of the conference. Association for Computational Linguistics. Meeting* (Vol. 2019, p. 6558). NIH Public Access.
- Liu, X., Shi, H., Chen, H., Yu, Z., Li, X. and Zhao, G., (2021). “imigue: An identity-free video dataset for micro-gesture understanding and emotion analysis.” In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10631-10642).
- Park, C.Y., Cha, N., Kang, S., Kim, A., Khandoker, A.H., Hadjileontiadis, L., Oh, A., Jeong, Y. and Lee, U., (2020). “K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations.” *Scientific Data*, 7(1), p.293.

Kosti, R., Alvarez, J.M., Recasens, A. and Lapedriza, A., (2019). "Context based emotion recognition using emotic dataset." *IEEE transactions on pattern analysis and machine intelligence*, 42(11), pp.2755-2766.

Stachl, C., Au, Q., Schoedel, R., Gosling, S.D., Harari, G.M., Buschek, D., Völkel, S.T., Schuwerk, T., Oldemeier, M., Ullmann, T. and Hussmann, H., (2020). "Predicting personality from patterns of behavior collected with smartphones." *Proceedings of the National Academy of Sciences*, 117(30), pp.17680-17687.