Auditory Affect Recognition

Project Report
Submitted in Partial Fulfilment of the Requirements
For the Award of the Degree
Of

Master of Technology
In
Signal Processing and Digital Design

by

Sudiksha Shukla
2K19/SPD/25

Under the supervision of

Prof. Rajiv Kapoor

Department of Electronics and Communication Engineering

Delhi Technological University

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

October 2021

Department of Electronics and Communication Engineering

Delhi Technological University

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Sudiksha Shukla, Roll No. 2K19/SPD/25 student of M.Tech (Signal Processing and Digital Design), hereby declare that the project titled "**Auditory Affect Recognition**" which is submitted by me to the Department of Electronics and Communication Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associate ship, Fellowship or other similar title or recognition.

Place: Delhi — Sudiksha Shukla

Date: 22/10/2021 — 2K19/SPD/25

Sudiksha

Department of Electronics and Communication Engineering

Delhi Technological University

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "Auditory Affect Recognition" which is submitted by Sudiksha Shukla, Roll No. 2K19/SPD/25, Electronics and Communication Engineering Department, Delhi Technological University, Delhi, in partial fulfilment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the students under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree to this University or elsewhere.

Place: Delhi **PROF. RAJIV KAPOOR**

Date: 22/10/2021 **SUPERVISOR**

ABSTRACT

The study of emotion recognition to lay groundwork for this work took a kick start by hovering through the so far conducted research driven by other researchers. According to a saying, one must look at the weather when they step out and observe people's faces when they step in, recognition of correct facial expressions (FEx) is a way to infer someone's inner emotions and intentions. However, assessing emotions just from FEx, unquestioningly could be a naïve lead as people may suppress their true emotions by articulating false FEx. Since, humans don't always wear their hearts on their sleeves, we can shift the source from FEx to voice or speech to get a better exactness of the true emotion. An emotion recognition system using speech (SER) makes a significant contribution in the field of computer vision in terms of assessing human behaviors using machines. In SERs, both accuracy and interpretability form equal value. In this paper, we aim at solving this issue by combining interpretability of Fuzzy rule based systems with accurateness of estimations done on the basis of Kernel Densities. The RAVDESS dataset has been used to analyze the performance by considering variants of possible combinations of the dataset entries.

ACKNOWLEDGEMENT

I would like to express my gratitude towards all the people who have contributed their precious time and effort to help me without whom it would not have been possible for me to understand and complete the project.

With profound sense of gratitude, I would like to thank Prof. Rajiv Kapoor, my Project guide, for his guidance, patience, support and encouragement throughout the period this work was carried out. His readiness for consultation at all times, his educative comments, his concern and assistance even with practical things have been invaluable.

Sudiksha Shukla
2K19/SPD/25

CONTENTS

Candidate's Declaration	**ii**
Certificate	**iii**
Abstract	**iv**
Acknowledgement	**v**
Table of Contents	**vi**
List of Tables	**viii**
List of Figures	**ix**
CHAPTER 1 – Introduction To Emotion Recognition Using Speech	**9-13**
CHAPTER 2 –Preliminaries	**14-25**
2.1 A system grounded on Fuzzy laws	
2.2 Density estimations using kernels	
2.3 Majority vote classifier	
CHAPTER 3 – Performance measurement	**26-28**
3.1 Measurement standards	
CHAPTER 4 – The proposed method	**29-39**
4.1 HF functioning	
4.2 DE functioning	
4.3 The HFF-DE method	
CHAPTER 5 – Implementation : Dataset Specification	**40-43**
CHAPTER 6 – Experimental results	**44-50**
CHAPTER 7 –Ground laying for future work	**51-58**
7.1 MEx dataset	
7.2 Literature review on MEx	
7.3 Challenges around MEx	
CHAPTER 8 – Conclusion & Future Work	**59**
References	**60-70**
List of Publications	**71**

List of Tables

Table No.	Title	Page No.
Table 5.1	Emotions covered in RD	41
Table 5.2	RD naming-convention	42
Table 5.3	RD durations	43
Table 6.1	Features extracted for an example of Neutral emotion	45
Table 6.2	Accuracies of bidding processes on RD	50
Table 7.1.1	Spontaneous and non-spontaneous datasets brought together	54
Table 7.2.1	MEx publications approximated frequencies	55
Table 7.2.2	MEx publications 2009-2020	55

List of Figures

Figure No.	Title	Page No.
Fig. 1.1	Facial expressions of 7 universal emotions	10
Fig. 2.1	A strong triangular-fuzzy-partition having 3 fuzzy sets	18
Fig. 2.2.1(a)	From HG to Density estimation – step 1	21
Fig. 2.2.1(b)	From HG to Density estimation – step 2	21
Fig. 2.1.1(c)	From HG to Density estimation – step 3	22
Fig. 4.1.1	Block diagram of Block A of proposed technique	35
Fig. 4.2.1	Block diagram of Block B of proposed technique	36
Fig. 4.3.1	FINAL BLOCK DIAGRAM : Blend of Block A and Block B of proposed technique	38
Fig. 5.1	Important features of RD	43
Fig. 6.1	Results for SVM classification	46
Fig. 6.2	Pareto font for test and TrS data for MEFs	46
Fig. 6.3	KnwB or rule-base taken up	47
Fig. 6.4	The triangle-membership functions for first few features taken up	48
Fig. 6.5	DE performance individually	49

CHAPTER 1

INTRODUCTION TO EMOTION RECOGNITION USING SPEECH

Emotions are habitually allied with motivation, temperament, mood, personality and disposition. One way to define emotions is to say that it is a robust feeling that arises from a state of mind which depends on situations Sentiments have a potential to be accepted as reactions sourced by external occurrences. The physiology of emotion is believed to be closely related with kick starting nervous-system, having varying situations and penetrations depending on unique sentiments. Also, emotions can be linked with behavioral-tendencies. The restrained individuals hide ones genuine sentiments. Emotions are frequently the momentum that drives buoyant / harmful gusto. Through another perspective, one’s sentiment is a diverse type which encompasses a wide variety of compelling psychological events. They might possess conspicuous tangible insertions or possess indistinguishability for bystanders. Their intensity might be increased, in other words, it becomes critical for a living or concentration of optimal well-being, otherwise can be inconsequential. A sentiment can be in a societal context appropriate or inappropriate.

FEx are one of the supremely imperative traits in interaction. A face may be described using the movements/ positions that the muscles hold, under ones facial sheath. FEx may be identified through non-vocal interaction. Communicating through emotions can be thought to be intriguing because it seems that some of the appearances (embarrassment, joy, anger, contempt, disgust, interest, surprise, sadness, grief, fear and shame) are believed to have been biologically-rooted, across societies. The above theory lays its ground on judgements which fixates on the fact that any FEx can be thought of as an artefact in terms of culture and social knowledge. FEx may be espoused both, unreservedly or compulsorily, over whom the approaches around neural-responsibility for overriding expression vacillate. Premeditated Ex generally are publically familiarized. In addition, the FEx have a strong reference to a particular psychology. Many psychology-experts are capable of perceiving ambiguous denotation by acquiring information through one's facial-attributes. FExs reveal colossal responsibility in the conduction of language. Countless language sayings incorporate FEx on screen.

Study of FExs have had impact on the history of research since the 1970s. FExs are one of the primary aspects of nonverbal communication among human beings. The face not only communicates thoughts but also internal emotions. Different FEx can be thought of different positions of the facial muscles or in other words, they may be calculated or unpremeditated movements, transpiring as a few or the whole shebang of 43-facial muscles engross. The modern theory on basic emotions [1] [2] [3] contributed to the boom of the research journey. It is generally accepted in the scientific convergence that 7 fundamental emotions exist, for these in particular, there is a logical confirmation that these FExs are in fact represented in a similar way by people of all cultures. They are also believed to be genetically encoded. They are anger, disgust, repulsion, fear, joy, pain and astonishment. These emotions can be stated basic emotions as they can be naturally detected by a human neuron system in our day-to-day endeavors.

Emotion identification logic is employed in innumerable applications such as driving logics in automobiles, for inferring mental distress, detecting deceptiveness, etc. To estimates emotions, hominids have informants like FRs or visual communication, temperament of voice, utterings, etc. Above mentioned are incorporated in innumerable appliances, including but not limited to, mortal personal computer (PC) collaboration, client sentiment exploration, gaming, and lots.

Fig. 1.1 Facial expressions of 7 universal emotions

Notwithstanding the nice evolvement within pitch of apperception, natural man PC interface persists to be thought-provoking as the PC doesn't comprehend the essence of the reciter. That has familiarised a fresh investigation pitch, namely SER, which suggests acknowledging spirit of reciter.

Resolving chi of one's intellect is on the lines of being tricky. As hominid -PC collaboration frolics a substantial character whilst this epoch, comprehending the humanoid emotions would lead to flawless harmonization between kindred and PC. Unambiguously, in VR, the psychosomatic assessment of captivation inside the VR atmosphere is basically the furthermost imperative stride. This captivation is priced on idea of relationship amongst emotive gist from dialog and consequently the contextual milieu.

The emotion identification using FEx forms a crucial field of study due to one challenge that it poses, which is the fact that FEx can be controlled and subsided by the being. This could be done calculatedly or sub consciously. In either cases, figuring out the true sentiment can be a rigorous process. Here, speech or voice of the subject under observation can contribute with a major weightage in revealing one’s true sentiments. In current paper, our focus would be on building a robust SER system which can in turn be made use of to reveal the true emotions of the subject with utmost accuracy and also focusses on being interpretable to the researchers.

Sometimes it is a challenge to identify the speaker's true feelings through his or her speech. Breathing, phonation, enunciation, and reverberation form the four primary mainstays of speaking practice . The sound-energy is driven by air flow. Noise is nothing more than trembling of vocal cords that occurs as the phonation progresses. Articulation is the process of sound composition. The silhouette and proportions of the pharynx, nasal and verbalized cavities contribute to the formation of resonance. Humanoids can use speech to diagnose the sensation of others. The human sensory structure shows an important function during this progression. Cepstral evaluations and statistical methods can be examples of this humanoid sensory structure. Well-known spectral scanning practices such as MFCC (Mel freq) and LFCC (Linear freq) are used in the SER system.

It is well known that classification systems based on the Fuzzy Language Rule (FLR) [4] are effectual dialectal models that are employed to decipher the complexity of cataloguing on the basis of linguistic-variable [5] [6] which are not only correct, but also interpretable. This type of linguistic model is easier to understand for seekers because of linguistic variables compared to numeric level models. Fuzzy rules are therefore widespread in real applications [7] [8] [9] [10]. However, the accuracy of FLR can decrease in solving some complex applications because the notion of phonological variables is not adequately bendable [11]. To dig a little deeper, the fuzzy-sets concept forms base principle of FLR, that is known to get hold of the fuzziness uncertainty but is unable to meritoriously address inexact or unfinished ambiguity [12]. For complex real-applications, however, distinct kinds of vagueness can exist, like incompleteness, inaccuracy and fuzziness [14] [15] [16]. Dealing with precarious information and making decisions in real-world applications is still an open problem [13] [17] [18]. Whereas, to review the undiscovered DE, KE [19] can be assumed as a tactic with a non-parametric basis, and it has been seen that outcome can be made functional to work around ambiguous issues [11].

Hence, in this paper, a fresh technique has been projected by enhancing the creation of a fusion (F) of a hybrid fuzzy concept (HF) that is based on fuzzy rule based classification system along with density estimation (DE) technique which in turn utilizes a base of naïve Bayes classifier as a building block. The fusion is established incorporating a model of voting classifier to bring out the decision formulating performance, which we would refer as HFF-DE. This implementation has been done particularly in the field of Auditory Affect Recognition. The DE aims at enhancing the flexibility of the system whereas; the HF caters to the issue of uncertainty. Therefore, utilizing the framework of voting classifier, the 2 techniques have been meticulously combined by maintaining the spot light on their corresponding pluses to broaden the improved accuracy and sturdiness. In the scaffold of voting classifier, 2 or more norms are treated as evidence coming through distinct sources of information in such a way that their outcomes that are evaluated as probability delegation are merged by means of majority voting of individual evidences. For aforementioned stratagem, concluding-class is established by largest assigned-probability.

The major milestones covered by this paper are given below:

- One fresh scheme is projected through amalgamating DE along with HF technique in skeleton around voting classifier, coined as HFF–DE in the auditory affect emotion recognition discipline.

- The HFF–DE scheme can deliver immense robustness and exceptional flexibility when uniting information with multi-feature.

- The HFF–DE classification system enriches the accuracy that has been tested on real dataset.

The work has been laid out in the given manner. Starts with Ch-2 that introduces briefly the elementary theories around HF, DE and MVC. Ch-3 then briefs about the performance measurements. How the projected HFF–DE system has been assembled by joining the dots between the HF and DE is conferred within Ch-4. Ch-5 argues the dataset incorporated on which the classification issue has been resolved along with the results in Ch-6. Ultimately, the conclusion is pulled in Ch-8, before which we lay the ground work for MEx recognition for our future work in Ch-7.

CHAPTER 2

PRELIMINARIES

For the physical realm, information from multiple sources can be inaccurate, uncertain, or even contradicting itself [20][21][33]. Dealing with unsafe information is important; therefore, many approaches are presented in [27][28] that are used in decision making [22][23], reliability assessment [25][26], industrial alarm systems [31][32], fault diagnosis [29][30], and recognition [24][11]. In this section, the aim is to create a basic understanding around the key concepts of the proposed method.

2.1 A SYSTEM GROUNDED ON FUZZY LAWS

We have observed from former studies, an information -instance may be portrayed as assembly constituting components in accordance to respective analogous operational affiliation, moreover above components largely consist the sets grounded on linguistic fuzzy aspects [34][35]. Contrasting to numeric-values, above dialectal tags enable operators with the ability to-create a superior understanding around the configuration alongside the triggering layer underneath primary records. In case primary records are in configuration hovering around fuzzy-sets, also instance of semantic evidence can be subjected as the most apposite form of exemplification. The splitting of features in accordance to function of fuzzy membership is the base to the fuzzy laws. Different situations may incorporate distinct fuzzy-laws. On lines of, the fuzzy selection of dichotomies, application-conundrum with 3 or even more characteristics.

This section focuses on creating a strong base around the key notions involved in a System based on laws abide by Fuzzy concept (FS). A system centered around fuzzy laws or FS is governed on the basis of 2 poles. Those 2 poles can be recognized by the notations, the Knowledge Base (or KnwB) and the Engine grounded on Fuzzy-Inference.

To describe the rules, there is a need of fuzzy sets which are in turn defined using certain parameters and an arrangement of linguistic rules. These rules and parameters together form the KnwB. The subjects in KnwB are then utilized in order to generate a prediction of a classification problem, for brand new inputs.

Firstly, laying grounds and declaring some variables and sets. Let the set of input attributes be :

$$Y = \{Y_{1,} \ldots , Y_H\}$$

and for the output attribute, Y_{H+1}.

For $h = 1, \ldots , H+1$,

let V_h represent the universe of the h^{th} attribute Y_h.

To act as a fuzzy partition,

let $Q_h = \{B_{h,1}, \ldots , B_{h,Sh}\}$

where, S_h be the arrangements grounded on features from fuzzy concept Y_h.

Now, the goal is to outline the data used for training (TrS) as :

$\{(y_1 , y_{H+1,1}), \ldots , (y_M , y_{H+1,M})\}$ as a pool of M input as well as output duos.

Also, $y_s = [y_{s,1} . . . , y_{s,H}] \epsilon R^H$, $s = 1, \ldots , M$.

Beyond this the path can be divided into 2 parts, namely, regression and classification depending upon the problem we aim at resolving. In problems that require regression for solution,

for all $s \epsilon [0 \ldots M]$, we have $y_{H+1,s} \epsilon R$ as we take Y_{H+1} to be a continuous attribute.

For N linguistic fuzzy rules, we tend to assume a Regressor grounded on the laws around Fuzzy-concept (FR) constituting of a rule-base of M regulations. The intention here is to approximate an output when given an input vector. Equation 1 gives a rule where,

$k_{n,h} \epsilon [1, S_h]$, $h = 1, \ldots , H+1$ is used to identify the manifestation of fuzzy-set, amongst S_h dialectal instances of Q_h compartment, that has been nominated for Y_h in rule D_n.

D_n : If Y_1 is $B_{1, jn,1}$ and … and Y_h is $B_{h,jn,h}$ and .. and Y_h is $B_{H,jn,H}$ then Y_{H+1} is $B_{H+1,jn,H+1}$. …. (1)

Whereas, for classification required problems Y_{H+1} is taken to be categorical and $y_{H+1,s}$ ϵ E. Here, $C = \{ C_1, \ldots , C_L \}$ would be the set of L possible classes. We can assume a Classifier grounded to the laws revolving around Fuzzy (FC) constituting of a rule base of M regulations. The intention here is to approximate an output when given an input vector. Equation 2 refers to such a rule.

D_n : If Y_1 is $B_{1, jn,1}$ and … and Y_h is $B_{h,jn,h}$ and .. and Y_h is $B_{H,jn,H}$ then Y_{H+1} is C_{jn} with W_n . …. (2)

Here, W_n represents the weight rule which illustrates conviction degree of the categorization falling to C_{jn} class **f**or a design to be in the right place within fuzzy sub-devision demarcated by the forerunner to law D_n [36]. The concept of weight W_n linking to a law has been portrayed with distinct point-of-views in collected works [37]. Where, C_{jn} is the class label corresponding to n^{th} rule.

For an unknown input say, $y' \epsilon R^H$ the FR or FC provides the anticipated output estimate or class label, correspondingly. The above stated is accomplished by embracing a particular inference engine. There exists various inference engines [37] but in both the above instances, the dilution around instigation for every law alongside the recording affects given output accordingly.

For our purpose, we would be focusing on triangular fuzzy partitions. The tuples ($a_{h,j}$, $b_{h,j}$, $ch_{,j}$) are used to represent the membership function which in turn can be associated to triangular fuzzy sets, say $B_{h,j}$ that form a partition. $a_{h,j}$ and $c_{h,j}$ are associated along left-ward and right-ward boundaries around $B_{h,j}$ as support followed by $b_{h,j}$ corresponding to principal position. Additional versions revolving around FS, to name some, TSK-FS, FS amid DNF laws plus FS grounded upon concept of manifold granularities, as viewed through other concentrated literatures [38].

Here, penalized certainty factor is utilized to evaluate W_n. Equation 3 gives the PCF equivalence.

$$PCFn = \frac{\sum_{Yh \in Class\ Cj} \mu Bj(Yh) - \sum_{Yh \notin Class\ Cj} \mu Bj(Yh)}{\sum_{h=1}^{H} \mu Bj(Yh)} \quad \text{... (3)}$$

We can evaluate the Matching Degree by calculating Y_h using a product or minimum T-norm. Equation 4 gives the above expression.

$$\mu_{Bj}(Y_h) = T(\mu_{Bj1}(Y_{h1}), \dots, \mu_{Bjm}(Y_{hm})) \quad \text{… (4)}$$

The decision function F is used for classification. Equation 5 depicts the same for :
s = 1, … , H+(1).

$$F(Y_1, \dots, Y_{H+1}) = \arg\max \{Y_s\} \quad \text{… (5)}$$

For an aggregation function f, soundness degree can be defined as equation 6 for s = 1, … , H+1.

$$(Y_s) = f\left(b_j^s, j = 1, \dots, L, and\ b_j^s > 0\right) \quad \text{… (6)}$$

Finally, the association degree is depicted by b_j^s as given in equation 7.

$$b_j^s = \{\ p(\mu_{Bj}(Y_h), W_j, j = 1, \dots, L\ if\ s = Class(D_j) \quad \text{...(7)}$$

and $\qquad b_j^s = 0\ , otherwise$

where, p → product T- norm.

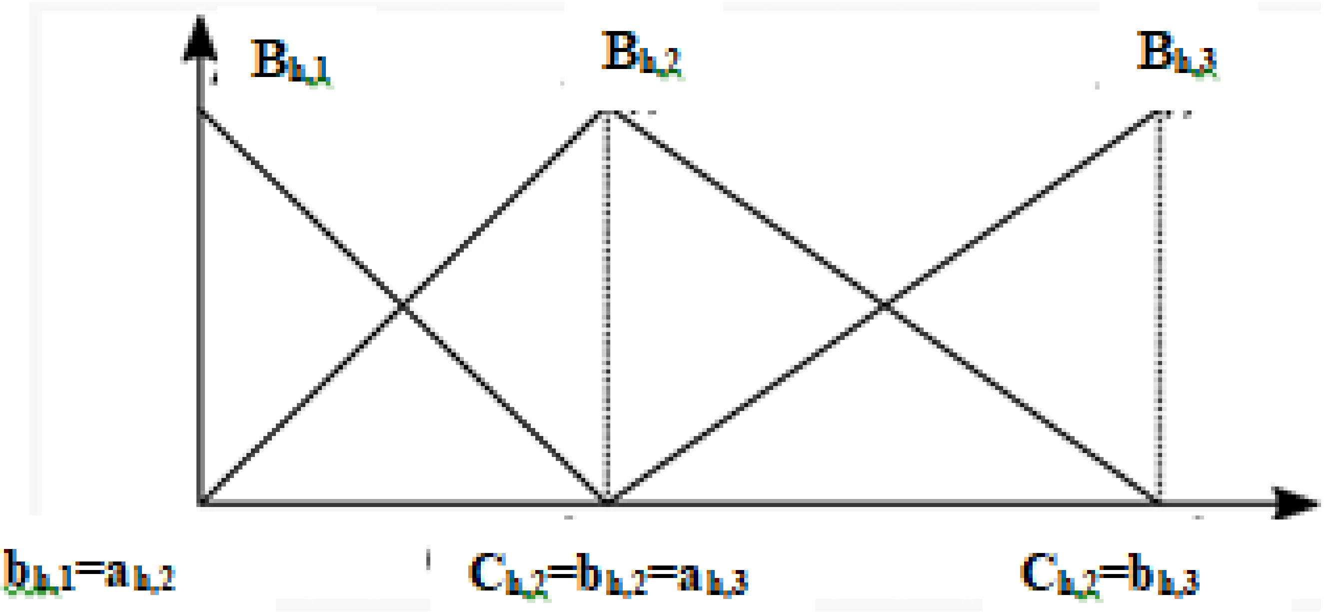


Fig. 2.1.1 A strong triangular-fuzzy-partition having 3 fuzzy sets

In order to create an FS, one needs to focus on 2 things. Firstly, establish the most suitable set of laws that can cater to the particular problem of classification or regression. Secondly, shift the focus to optimal amount of fuzzy-sets required around every feature along with respective considerations. The purpose here is to model the interpretability and exploit the accuracy. If we focus on lessening the estimation error of the output instances, we can increase the accuracy of an FR. Whereas, for our case where we plan to aim at a classification problem, FC accuracy is dependent on the number of patterns classified correctly. When it comes to explainability for an FS, the focus is on explaining competences of answering the whys and hows around the predictions done and making it easy to comprehend for humans, coined as interpretability. Thus, most concerns revolve around the simplicity of inference engine that would conclude from laws and facts. The ability of comprehending the construction of the model, also known as transparency also plays a crucial role here. If the fuzzy partitions keep a decent integrity and the linguistic RB contains lower quantity of laws and clauses, then the FS can be branded as having an elevated transparency level. The veracity associated with fuzzy-partitions hinge on various characteristics, namely, distinguishability, order, normality and reportage [37]. Researches have provided various measures for assessing FS interpretability, considering intricacy and semantic outlooks for RB and partitions.

2.2 DENSITY ESTIMATIONS USING KERNELS

Density estimations made with the help of kernels or KE acknowledged as one of many widely incorporated non-parametric methods [41], especially in the field of statistics. Its forte lies with data bases that do not let specification bias affect them, providing heights of flexibility in modelling such data. Any predetermined attributes do not persuade the estimations, thus forming a vital key trait and the name non-parametric.

A density estimator (DE) is a set of rules that processes a set of data with dimension-D and spawns a guesstimate around probability-distribution with same dimensions from where the information was extracted originally. Some algorithms achieve the above stated by using weighted sum of Gaussian distributions to symbolize the density. In a sense, KE is an algorithm that puts the Gaussian mixture concept to the logical ultimate state. A blend is utilized which is made up of a single Gaussian-component for every point, occasioning towards a primarily non-parametric density-estimator. Equation 8 gives a mathematical representation of KE for a univariate independent TrS data sample :

$$v = (v_1, v_2, \ldots , v_n).$$

$$f'_s(v) = \frac{1}{n} \sum_{i=1}^{n} K_s(v - v_i)$$

$$f'_s(v) = \frac{1}{ns} \sum_{i=1}^{n} K \frac{(v - v_i)}{s} \ , \ v \epsilon R \qquad \ldots (8)$$

Here, K → Kernel function and s → smoothing parameter or bandwidth (BW) and this is used to obtain a decent estimate of the density.

For one-dimensional data, a well-known simple DE is the histogram (HG). An HG splits the input information within disconnected tubs, reckonings the amount of instances present for every chest, hence exhibiting the product intuitively.

In case of a normalized HG, the height of the bins reflects probability density instead of reflecting counts. The downside of having HG as a DE is that representations with quantitatively distinct attributes can be attained depending on adoption of location and size of bin. In order to cater to this issue we can refer to the below mentioned steps. Figure 2.2.1 provides the pictorial depiction :

- Imagine HG to constitute of stack of blocks such that we stack 1 block inside of a particular bin on top of every instance in the database. This leads to a misalignment between instants and respective blocks.

- To solve the above issue, we shall stack the blocks associated with the instants that they represent, instead of aligning according to the bins.

- Adding the influences of each block at every location along the bin axis.

- The rough edges do not reflect any true information about the data.

- Replace the blocks present at every position by means of a smooth function say, Gaussian. Thus, using standard normal curvature at every instant in place of a block.

- The smoothed result obtained using a Gaussian distribution gives comparatively more accurate conclusions.

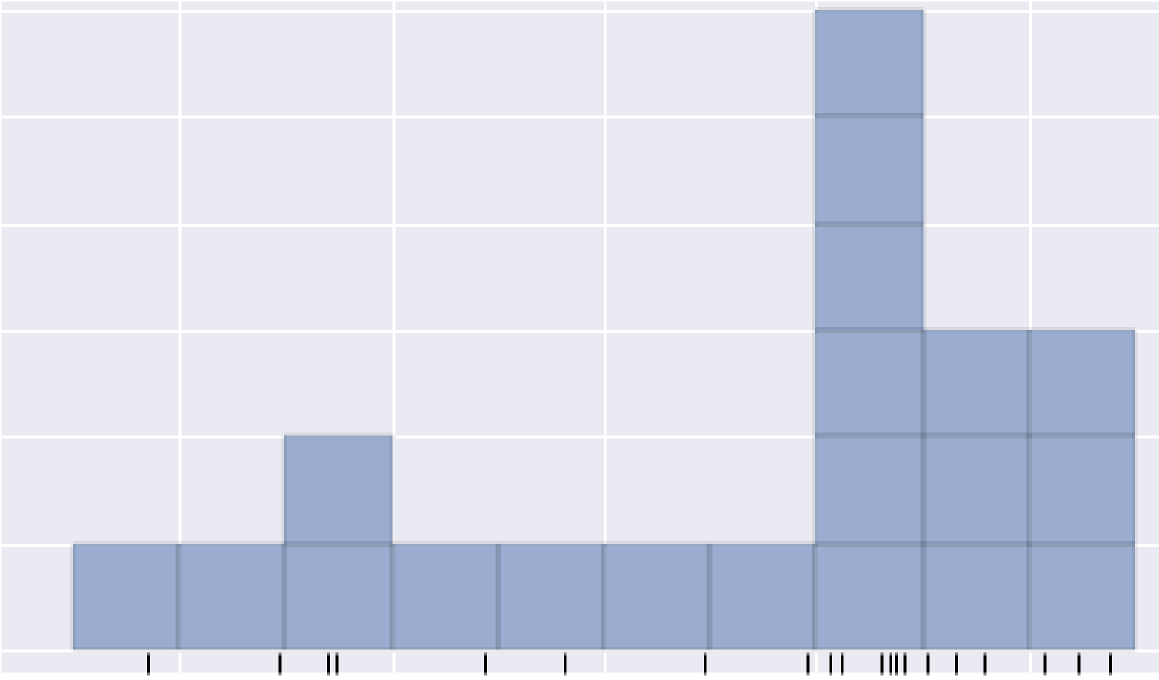


Fig. 2.2.1 (a) From HG to Density estimation

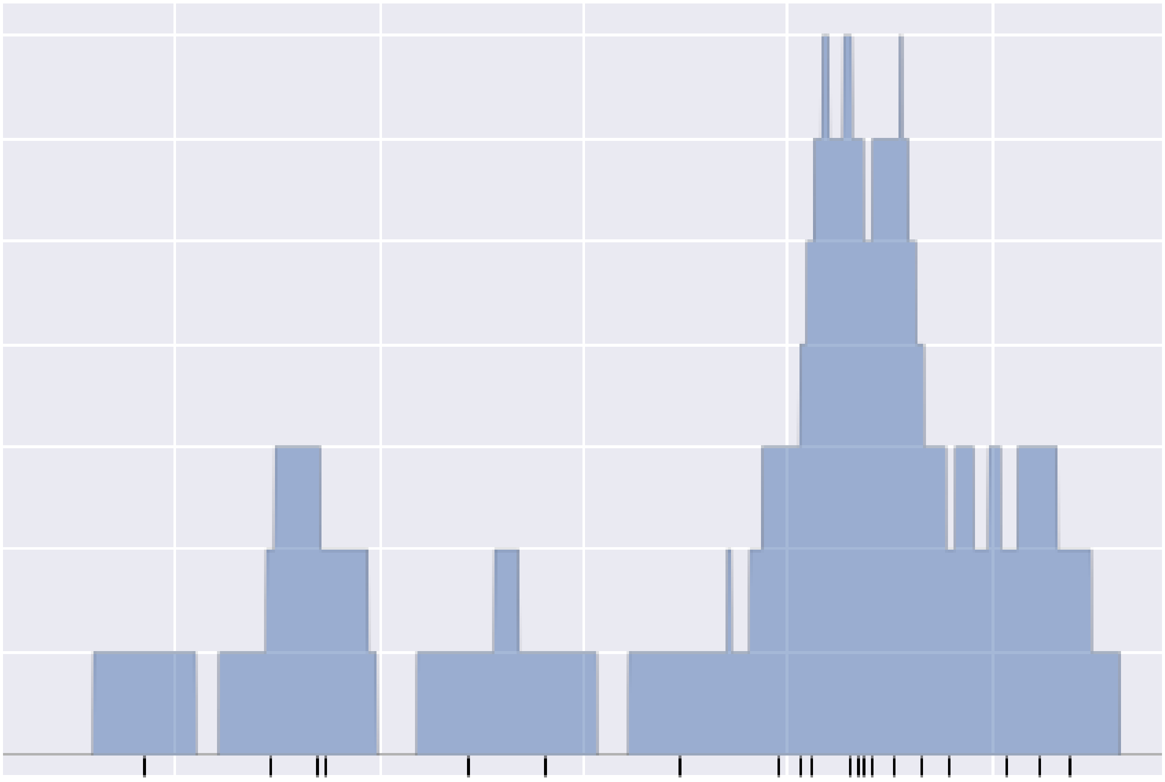


Fig. 2.2.1 (b) From HG to Density estimation

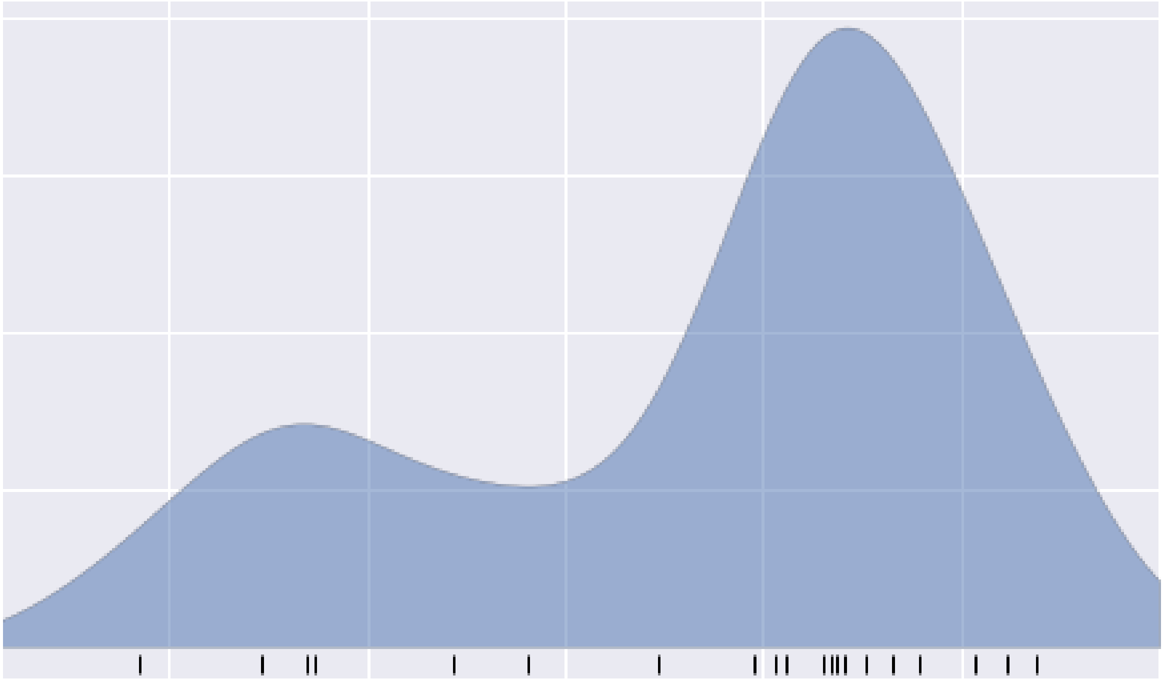


Fig. 2.2.1 (c) From HG to Density estimation

Fig 2.2.1(c) depicts an example of KE using a Gaussian kernel. As discussed above, the key attributes of KE is choosing the appropriate kernel that would specify the shape of distribution at each instant. The other is the BW that would determine the size of the kernel at each instant and acts as a knob to moderate the variance-bias trade off. A tight BW would lead to an estimate with a heighted variance (over fitting), in this case even a singular instant could lead to huge changes. A loose BW causes an estimate with a heighted bias (under fitting), in which case due to wide kernel the structure of the data is swallowed.

2.3 MAJORITY VOTE CLASSIFIER

After a classification model has been created, a brand new data instance can be given to the algorithm which then attributes an appropriate label to the instance basing its decision on the inputs of the variables. If various distinct classification techniques were to make choices basing the decision upon data instances' inputs, possibility of varying solutions exists. In such cases, a voting classifier (VC) comes into picture. A VC is a categorization technique that incorporates manifold classifiers to produce the prediction, ensuring high accuracy. It is mostly applied to either combine different fortes of various techniques and create an enhanced method or to resolve situations of confusions. The decision making is done on the basis of most frequent one.

VC merely aggregates the outcomes of every classifier tossed to the VC and makes the decision of which output class is to be selected on the basis of the most elevated majority of electing. The concept is to produce a one-stop model that is trained through several models and predicts the outcome on the basis of their collective majority of electing for every class, instead of producing discrete devoted models for each one of them. Two kinds of votings are reinforced by VC namely, hard and soft voting.

When the prediction class is the one with largest majority of votes or the one that had largest probability of prediction by every classifier, it is known as hard voting. For example, 3 models predict outcome class as [P, P, Q], then P would be the chosen outcome. In cases where, the final class prediction is done on the basis of mean of probability assigned to that class.

For example, for 3 models the probability for P = [0.53, 0.3, 0.47] and Q = [0.40, 0.2, 0.32]. Mean for each is P = 0.4333 and Q = 0.3067. Thus, class P wins.

Assume a classification problem where we have duos of observations (a_i, b_i) for $i = 1, \dots, n$.

We aim to use these duos to create a prediction law for B given A. We assume B is a categorical random variable instead of having $B \subset R$ or B being ordinal. The above statement implies that B does not have any ordering but actually denotes a class that an instance falls into.

Now, for 3 distinct classification rules say, g1(A), g2(A) and g3(A) we need to create a classifier which would be more enhanced than any of the individuals. One such way would be to pick the class that obtains greatest amounts of classifications or votes for every instance of A. Equation 9, gives the mathematical representation of a majority VC.

$$C(A) = mode\ \{g1(A), \mathrm{g2(A)}, \mathrm{g3(A)}\} \qquad \dots (9)$$

To portray the improvement which may be achieved using Eq 9 let us consider an example. Let the predictor space be distributed into 3 areas. In the first area, say g1 and g2 perform accurately but g3 given an inaccurate outcome. For the second area, g1 and g3 are precise whereas g2 imprecise and finally, the last area, g2 is correct along with g3 leaving g1 to be incorrect. If a test instance has an equal probability to be in any of these areas, each of the classifier would be inaccurate $1/3^{rd}$ of the time.

Whereas, Eq 9 would always lead to a correct solution.

An extension of the above is possible for any amounts of classifiers. We may also choose to give higher weightage to some particular classifiers.

Let g1, g2, … , gm be m laws from which votes have to be taken into account. Equation 10 then depicts the VC for such a case.

$$C(A) = \arg max_i \{ \sum_{j=1}^{m} wj\ I(gj(A) = i) \} \qquad \dots (10)$$

Here, w1, … , wm sum up to be equal to 1 and are the weights.

I( ) → indicator function.

For $w = 1/m$ we can get to eq 9 for m laws. As in our case, instead of dealing with classification rules provided by each classifier we may also incorporate the probability estimates produced by each classifier. This is shown in equation 11.

$$C(A) = \arg max_i \{ \sum_{j=1}^{m} wj \ \ P'_{i,j} \ \} \qquad \ldots (11)$$

Here, $P'_{i,j}$ → probability estimate for the i^{th} class j^{th} classification law.

For $P'_{i,j} = 0$ or $P'_{i,j} = 1$ the above formulation can be boiled down to eq 10.

In this section, we have thoroughly discussed the key concepts involved in the projected technique. Up-coming section aims to join given dots between all these concepts and describe the whole flow.

CHAPTER 3

PERFORMANCE MEASUREMENT

The progress in the field of SERs has been evolving drastically and thus to ensure that it is correctly directed a standard evaluation or measurement system is essential. This section focusses on the performance metrics along with the commonly used validation methods.

3.1 MEASUREMENT STANDARDS

Metrics is a system of measurement or a standard format which one could follow in order to measure their findings. SERs particularly incorporates measurements for the classification stage which is generally binary in nature. The main purpose is to compute the four parameters TN (True Negative), TP (True Positive), FN (False Negative) and FP (False Positive) recognitions. Most common basis used in the area of interest as performance metrics is accuracy. Accuracy is defined as the sum of True Positives and True Negatives divided by all the possibilities that are True Negatives, True Positives, False Negatives and False Positives. It is a percentage measure of how well the recognition process is working. The mathematical formulation is given in equation 12.

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \quad \dots (12)$$

Accuracy was the main performance metrics in the earlier studies, a few other standards were introduced as the different aspects evolved. Precision and Recall are two such standards. To quantify the exactness of the outcomes in fractional forms, precision shall be calculated. Once the cases have been rescued, among these the fraction of relevant cases contribute to the precision measure. Whereas, recall is associated to the fraction of relevant cases that can be rescued. It can also be referred to as sensitivity towards the results.

Since, both these metrics are allied to the significance of the outcome of the classified data, it is in best practice to utilize them together instead of accommodating them individually. Thus, for an exhaustive data investigation, another formulation can be made use of , instead of the basic accuracy. F- Measure, also known as F1-Score is calculated as the harmonic mean of the 2 parameters, namely, Precision and Recall.

Precision, Recall and F1-Score are depicted in equations 13, 14 and 15 respectively.

$$Recall = \frac{TP}{TP+FN} \quad \ldots (13)$$

$$Precision = \frac{TP}{TP+FP} \quad \ldots (14)$$

$$F1 - Score = \frac{2*TP}{2*TP+FN+FP} \quad \ldots (15)$$

Where, TP is True-positive, FN is False-negative and FP stands for False-positive.

As depicted from the above equations, FN or False-negative parameter has not been accommodated in this form of measure. Thus, creating ROC curves can be tedious.

In order to overcome the above issue, yet another standard evolved, called the Matthews Correlation Coefficient, abbreviated as MCC.

Along with the above parameters TN is also housed here. The output for MCC varies between -1 and 1. Here, -1 represents overall dissimilarity, 0 stands for an arbitrary guess and +1 symbolizes complete similarity.

Along with the merits of this being a better tool, comes the difficulty of not being able to evaluate all the detection types in every classification process. Equation 16 gives the calculation of the MC coefficients.

$$MCC = \frac{TN \times TP - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad \ldots (16)$$

CHAPTER 4

THE PROPOSED METHOD

An introduction to a fresh method called the HFF-DE that revolves around the HF and DE under the scaffold of voting classifier or MVC to cater to classification problems in the auditory affect emotion recognition field is provided in this segment. The block diagram of the technique has been divided into 3 parts for ease of understanding and representation.

DE is a strong tool for unraveling categorization problems due to it being non-parametric in nature and also data-driven which ensures a reduction in influence of uncertainty. Besides, fuzzy laws may deliver are interpretable develops that form a basis of linguistic variables. It can provide a comprehensible clarification of the final decisions in categorization problems. In order to further enhance the categorization exactness, 2 methodologies are processed collectively based on MVC.

4.1 HF FUNCTIONING

Most ML and AI models offer great accuracies but are weak on the transparency end i.e., the cause of the outcome is not very clear. Lately, the concept of explainable artificial intelligence (EaI) has gained popularity. EaI procedures and set of rules intend to drive AI-based systems and techniques towards higher transparency simultaneously sustaining the appropriate expected standards of precision and/or accuracy[42]. An FS is one such illustration that is angled towards explainability. Furthermore, the uncomplicatedness associated with the cognitive technique, incorporated in order to offer conclusion, grounded on input-facts, helps maintain an elevated explainability marks of FSs [43].

Recently, Evolutionary Algorithms revolving around Multi-Objective (MEAs) are being efficaciously incorporated in the process of producing FSs utilizing information, tending towards the so-called Multi-Objective-Evolutionary-Fuzzy-Systems (MEFSs) [40][38]. The MEFs are created keeping in mind to optimize explainability of FS.

There are certain software tools that produce EaIs and also analyze them. GUAJE [44] and ExpliClas [45] are 2 such tools utilized for designing interpretable models but they do not focus on optimizing explainability.

In the last few years, MEAs have been efficaciously implemented with keeping in mind to design FSs by concomitantly augmenting the explainability and accuracies of the system, which in turn led to the introduction of MEFSs [38]. Undeniably, MEAs enables one in establishing an approach to a process to augment a situation wherein 2 or more disagreeing objectives are aimed at being optimized concurrently, for our case those 2 objectives would be accuracy and explainability associated with FSs. MEAs gives collection of resolutions that are non-dominant and are branded by diverse trade-offs within chosen ideas, that portrays an estimation of Pareto-front [39]. By espousing learning scheme for Multi-Objective (MEL), it is likely to absorb outline of FSs utilizing variety of approaches, including but not limited to, acquiring only the KnwB reflecting on the fuzzy partitions that have been pre-defined, focusing on the fuzzy attributes only, plump for laws and settings, from a preliminary set of laws, and absorbing/picking laws synchronously with the enhancement of the fuzzy attributes. A comprehensive catalog of MEFSs could be looked for [40].

Usually, an MEL proposal involves an encryption corresponding to chromosome arrangements, which is generally dependent on the kind of FS, on the exclusive learning policy, and an arrangement of operators for mating, specifically crossover and mutation. These are fittingly outlined for functioning on the chromosome and spawning the next generation. Technically, an MEL policy must be governed by an exclusive MEA which would be responsible for managing the evolutionary process of multi-objective optimization. Whilst the ongoing of this course, a contender solution is assessed by deciphering the chromosome in order to form the concrete FS.

Precisely speaking, the accuracy of an FS is evaluated by embracing a given input and treating it as a TrS. Whereas, the explainability is calculated around a measure which is pre-defined, including but not limited to the quantity of rules or the sum amount of specifications in the KnwB, referred to as the Total rule Length (TL). Eventually, towards the conclusion part of the optimization, a group of FSs, spun around various trade-offs within interpretability and accuracy are reverted.

In our work, we aim at using an employment of a particular MEFS, known as Pareto archived evolution strategy – Rules and condition selection or PS-RC [46]. The PS-RC chooses a set of conditions and laws that are deduced from the original set and are fewer in amount, whilst the learning evolutionary procedure of multi-objective. To be completely accurate, here a fused version of PS-RC with a fuzzy decision tree algorithm has been acquired for implementation coined as, PS-RC-FT, whose main aim is to populate the initial set of laws [47].

A sub part of the above stated MEL is the (2+2)-M-PAES, which is an upgraded translation for popular (2+2)- PAES [48]. This is an evolutionary algorithm revolving around steady state multi-objective which accommodates 2 contemporary outcomes then archiving the non-dominant outcomes[49]. Whereas, given basic (2+2)- PAES preserves the outcomes only till the time they are substituted for the ones with specific attributes. This MEA has been implemented in some previous works [50] [51] [52] [53]. Here, a bi-dimensional vector is accounted for each chromosome. The vector's foremost element computes TL by computing the complexity of the system and the next term is utilized to analyze the accuracy with respect to the classification rate.

In Pc-RC, a chromosome say, Ch collates a solution for the system's issue. The Ch constitutes of (Ch_R , Ch_T). Where,

Ch_R → rule base representation and
Ch_T → representatives of the fuzzy set positions or cores.

Assuming K_{DT} to be the primary set of N_{DT} laws that have been extracted from the decision tree. To main the KnwB as interpretable we need it to be compact and hence the solution can at max hold N_{max} laws. The Ch_R codifies the KnwB as a vector with N_{max} pairs
$q_n = (s_n , u_n)$,
such that $k_n \in [\ 0, N_{DT}]$ is used to identify the law chosen from K_{DT}.

Also, a vector is needed which would be binary in nature and specifies if the condition is true in the chosen law, this is done for every feature Y_f, represented by $u_n = [u_{n,1} , \dots , u_{n,F}]$.

To be more particular, say $k_n = 0$ then the n^{th} law is not incorporated in the KnwB. Using the above procedure, we would be able to draw rule-bases containing reduced amount of laws than N_{max}. Also, say $u_{n,f}$ is equated to 0, the f^{th} clause of n^{th} law is swapped by a *do-not-care*.

Ch_T → vector constituting of F sub-vectors made up of $L_f - 2$ numbers ϵ *R*.

Here, an f^{th} vector say, $[c_{f,2}, \ldots, c_{f,Lf-1}]$ specifies the partition core's location. The partition veracity is warranted by the existence of robust fuzzy partitions. Thus, reportage, order, normality, distinguishability, etc are also warranted. With the aim of populating the next generation, mating operators are brought into picture, namely, mutation and cross-over. For mutation, 2 different tools for Ch_R and Ch_T are applied. To cater for the other, BLX-αα-cross-over with a certain α is employed on α Ch_T and for Ch_R, 1-point-cross-over is employed.

Fig, 4.1.1 depicts the block diagram of the working of our first block, say Block A, which includes applying the HF part of the proposed algorithm on the dataset. The following points, in mentioned order, highlights the process followed :

I. We start with splitting the given data into TrS and testing sets (TsS). Furthermore, we divide the TrS data by label.

II. On the split data, then the PS-RC algorithm is harnessed.

III. The above step provides a set of FSs that each carry a different trade-off within explainability and accuracy, that are expressed through TL and rate of classification, respectively.

IV. On the basis of above findings, the initial robust partitions (fuzzy). This is done by incorporating a fuzzy discretizer.

V. Then utilizing a multi-way decision-tree (fuzzy), preliminary set of laws are deduced.

VI. Every input feature is then partitioned within pre- outlined amount of sets (fuzzy) and preliminary laws are engendered.

VII. Model fit : list of FCs are spawned with the desired trade-offs.

VIII. An approximated Pareto font of FSs are thrown back, sequenced according to ascending accuracies.

IX. The values for the explainability and accuracy measures are extracted, after adopting the desired model from above.

X. The result is measured in Pareto font approximations on the TsS.

4.2 DE FUNCTIONING

In this paper, the aim was to incorporate Gaussian function into density estimation process and further combine the obtained procedure with the Naïve Bayesian classifier (or NBc). The NBc**,** is a course that involves formation of a primitive propagative model in respect to every class taken in the given conundrum. This is done to enhance the procreation rate of the system.

Here, our aim is to connate the Gaussian with the NBc to form the Gaussian Bayes (GNB). For GNB, the propagative model is formed by straightforward Gaussian aligned on an axis. Now, coming to the portion of forming an alliance between the GNB and KE, which is pronounced next.

By having a density-estimation system like DE, that forms an alliance between the KE and the GNB, we could eradicate the naive aspect from the picture. This would give us an opportunity to bring a higher level of sophistication to perform the same cataloguing with a more procreative model for every individual class. It would still be considered as Bayesian-classification, with deducting the element of “naïve”.

Fig, 4.2.1 illustrates the block diagram of the working of our second block, say Block B, which includes applying the DE outlook of the proposed algorithm on the dataset. The following points, in mentioned order, highlights the process followed. The algorithm is forthright and insightful to comprehend. The broad-spectrum methodology for generative-classification is given :

I. We start with splitting the given data into TrS and TsS. Furthermore, we divide the TrS data corresponding to label.

II. Corresponding to every set, mount a KE in order to acquire a generative-model of the input.

III. This would enable us to calculate a likeli-hood $P(A \mid B)$. Where, A is the observation corresponding to label B.

IV. According to the amount of instances belonging to every class in the TrS, we evaluate the class-prior say, $P(B)$.

V. Then for every unspecified instance say, a, the posterior probability corresponding to apiece class would be given as :

$$P(b \mid a) \propto P(a \mid b)\, P(b) \qquad \ldots (17)$$

Where, b is the corresponding label for a.

VI. Finally, the classification category that gives the maximum estimate for this posterior is chosen as the label for the instance and is assigned to it.

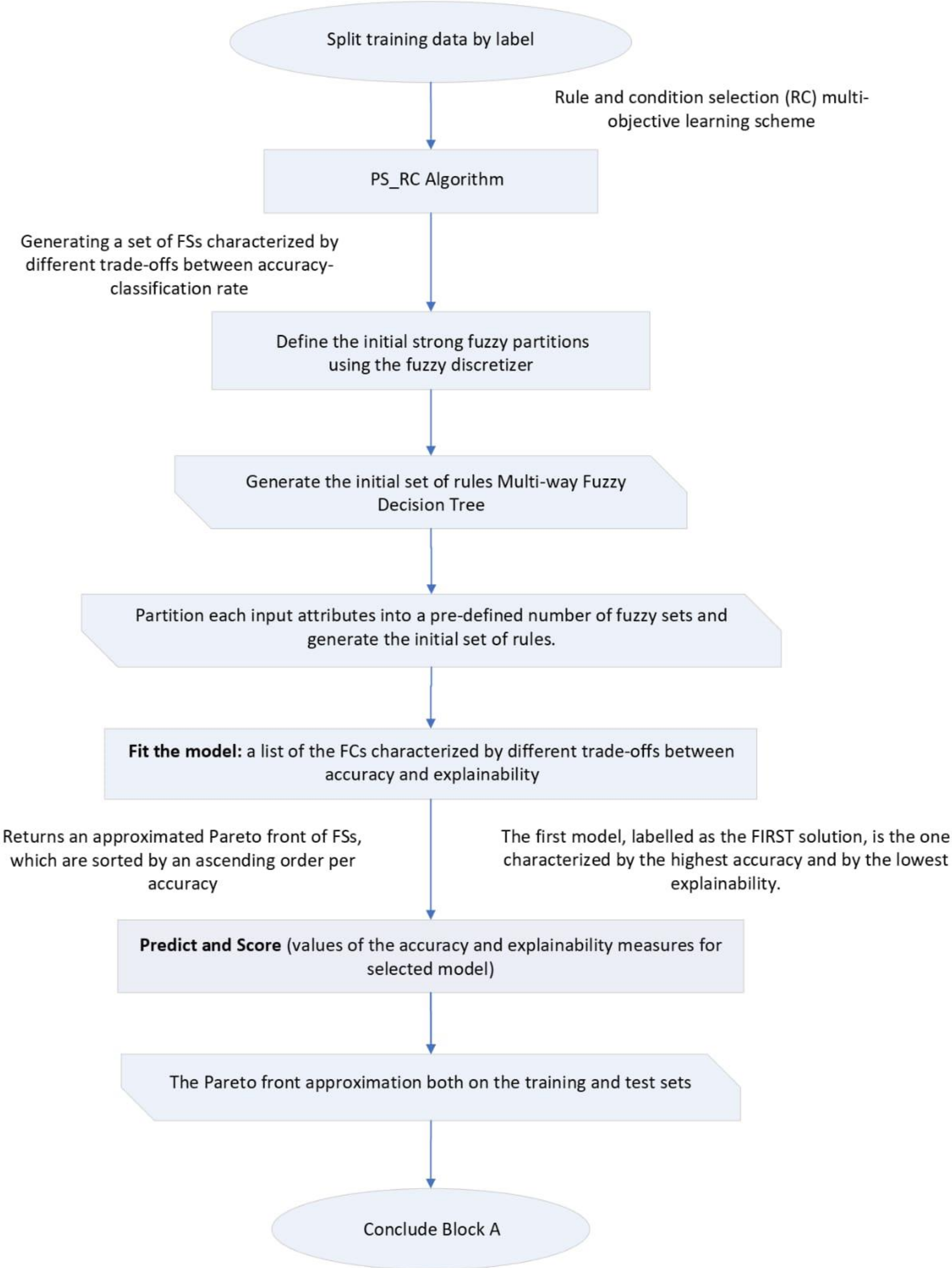


Fig.4.1.1. Block diagram of Block A of proposed technique

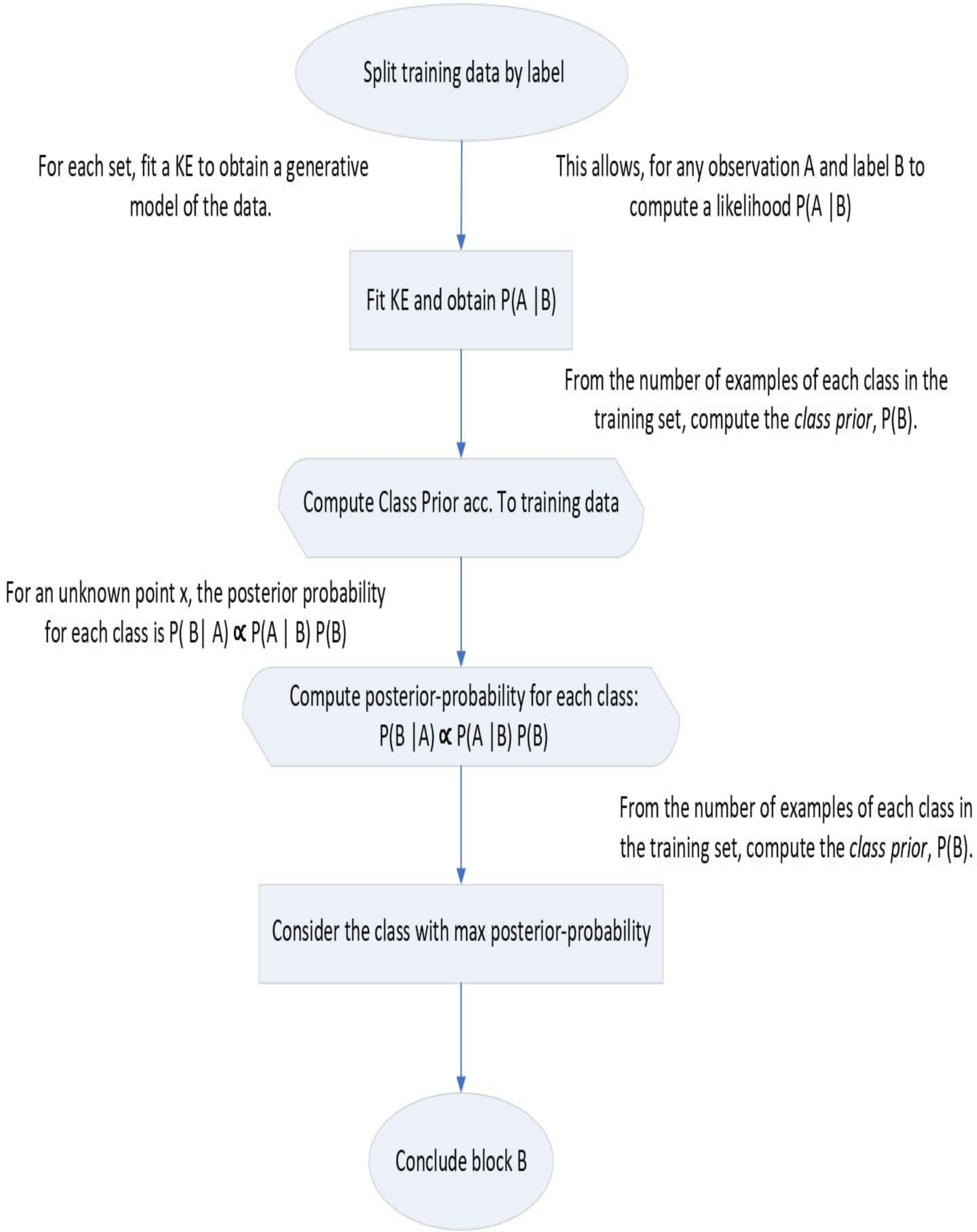


Fig. 4.2.1. Block diagram of Block B of proposed technique

4.3 THE HFF-DE METHOD

Finally, once we have successfully evaluated the 2 blocks on the data then we can go on and combine the results obtained from both. Now, we extract the classification-probabilities for both HF and DE and syndicate them using eq (18) as described above, after altering them tending towards pignistic-probability in order to form the concluding verdict.

$$p = p1 \oplus p2 \qquad \dots (18)$$

Here, p_1 → final probability extracted from HF and

p_2 → final probability extracted from DE.

p → concluding matching degree vector and the values symbolizes the support-degree of test-sample ϵ a particular class[11].

The concluding analysing function G that would come up with the label for the appropriate class along the lines of the maximum deduced from formula below:

$$C(p) = \arg max_{i=1,\dots N}\{\, p \,\} \qquad \dots (19)$$

A detailed description of this technique is depicted in Fig. 4.3.1 which combines Block A and Block B to form the overall technique.

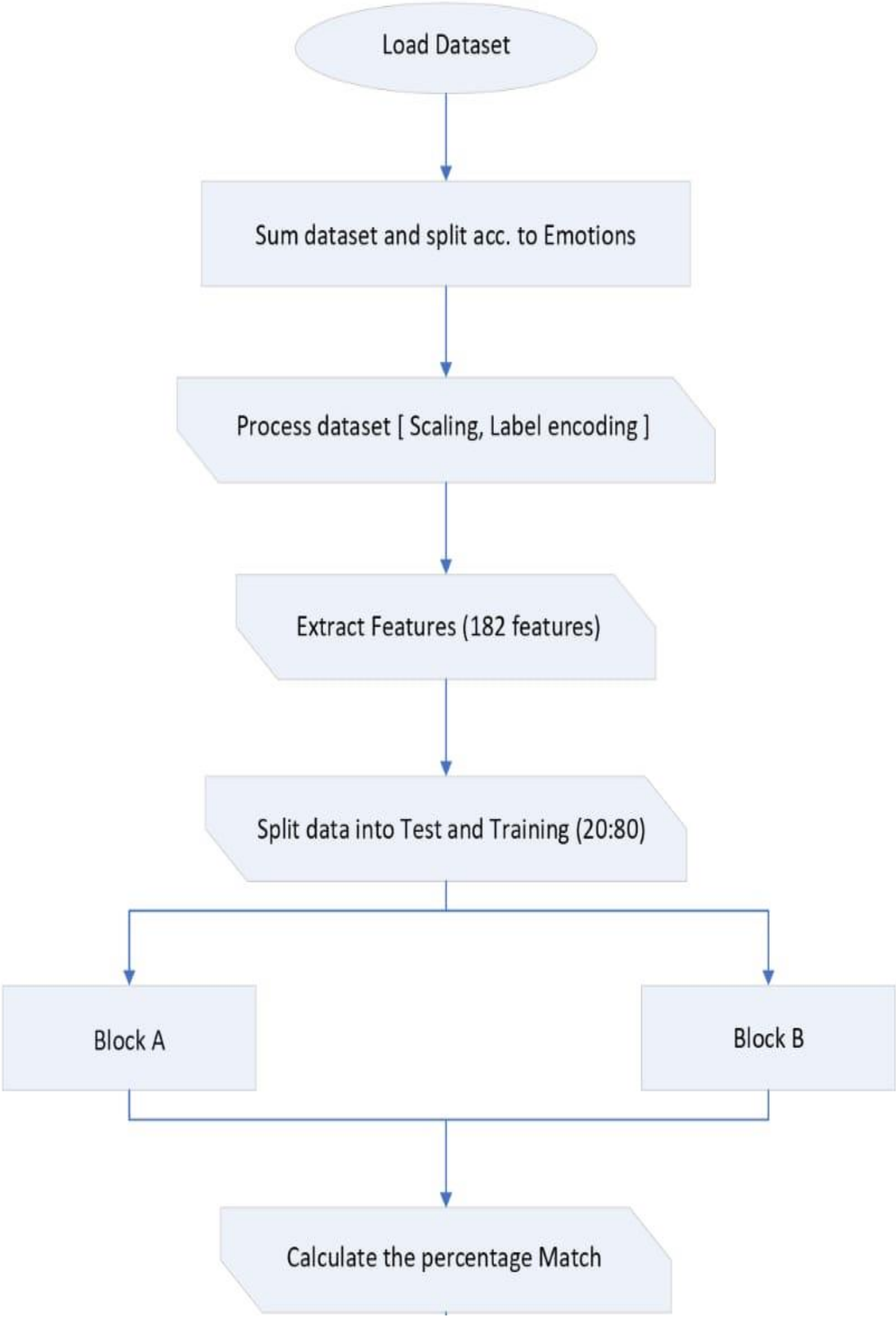


Fig. 4.3.1. FINAL BLOCK DIAGRAM : Blend of Block A and Block B of proposed technique

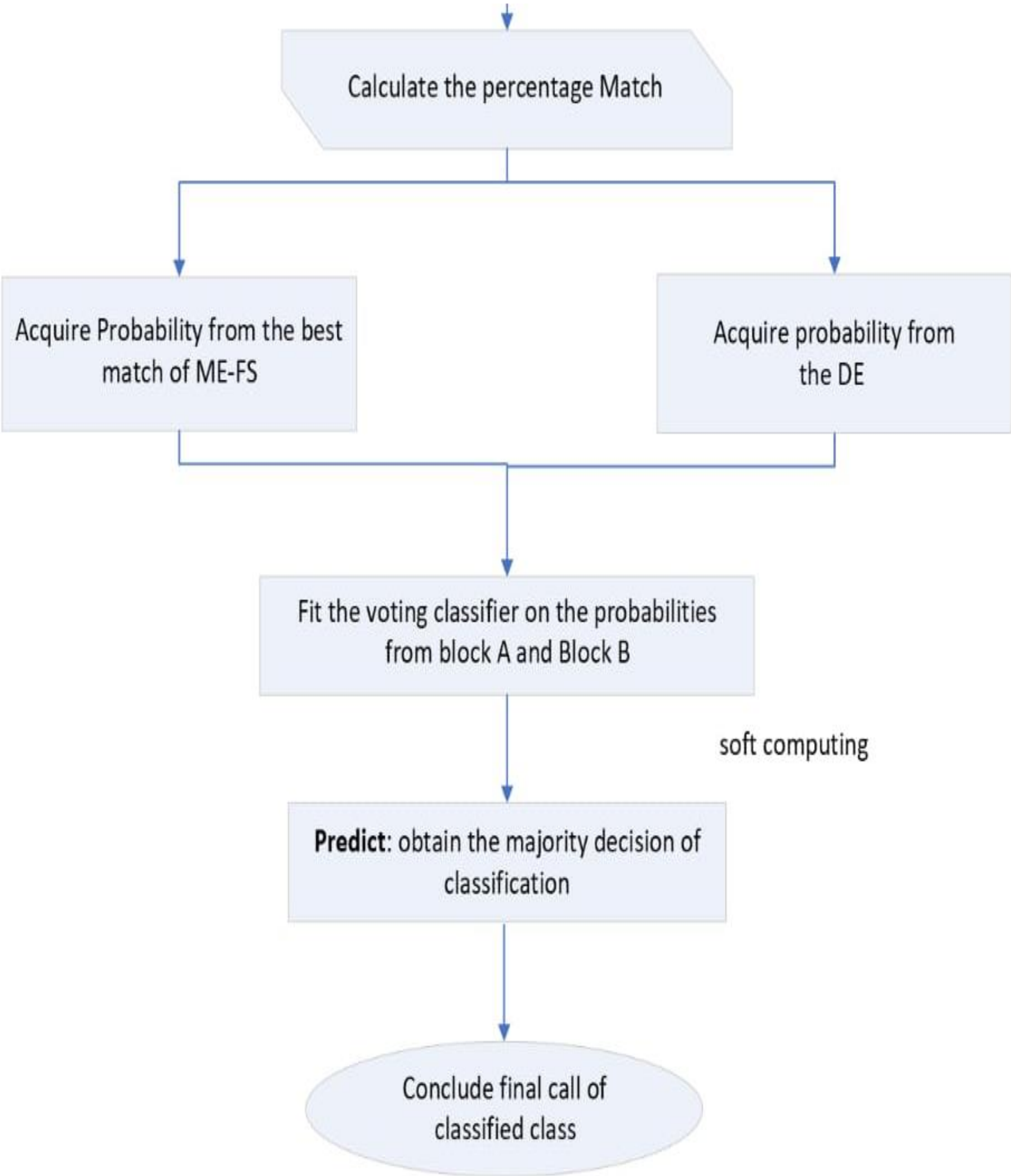


Fig. 4.3.1. FINAL BLOCK DIAGRAM : Blend of Block A and Block B of proposed technique (CONTINUED)

CHAPTER 5

IMPLEMENTATION : DATASET SPECIFICATIONS

The classification relevance has fascinated many researchers lately, as depicted in the works [54] [55] [56] [57]. To bring out the legitimacy of the approach, there has been a comparison done with the SVM, Naïve Bayes, MEFs or FSs and KE. The standard SVM based approach has been known for providing decent accuracy in the field. We aim at discontinuing the "naïve" aspect as a part of our approach thus a Naïve Bayes comparison formed a must existence for our work.

In this Segment, a brief peek into the details around the dataset used for the analysis of HFF-DE procedure is highlighted. The exercised data is the audio part of an audio-visual database. The RAVDESS dataset has been employed in this experiment. There are a few versions of datasets present for research work, namely, natural, acted and induced. These categories are introduced on the basis of the acquiring techniques involved.

- **Natural dataset** incorporates a few sum of emotions due to their nature of being private and would imply copy-right concerns. For them to be raw, greatest concern comes up of them being mild in existence and making it tricky for computers to catch them.
- Next are the **induced** ones, largely recognized as provoked as the emotions are generally captured by screening tapes and producing apt situations and hence are stimulated.
- The last one is the **acted set** where actors are asked to perform the category of emotion in consideration and are referred to as simulated. These are expected to be a lot more dramatic and this usually done by giving a neutral statement to be acted out with various emotions.

The Ryerson Audio-Visual Database of Emotional Speech and Song or RAVDESS (RD) was introduced by Russo and Livingstone[58]. The RD incorporates North-American English as language and picked actors. Twenty-four performers among who a 50-50 proportion of both genders was maintained were considered for recording. In total, 58 had interviewed. Figure 5.1, tabulates the key features around the RD revolving around amount of samples, lingual, script and participants. RD contains speech-signals that can be further distributed among different sentiments. Table 5.1, briefs about the distinct emotions incorporated in RD. These are included in 2 forms depending on the concentration modulated by the voice by artists.

EMOTIONS in RD	Normal Intensity	Strong Intensity
Angry	✔	✔
Calm	✔	✔
Happy	✔	✔
Sad	✔	✔
Surprise	✔	✔
Fearful	✔	✔
Disgust	✔	✔

Table 5.1..Emotions covered in RD

Along with the below mentioned, an auxiliary neutral emotion is also added in the RD. In order to work with this data, one needs to familiarize themselves with the naming-convention followed by the RD. Every one of the 1,440 records is identified by an exclusive handle. The file-handle involves a seven-part identifier suffixed by wav extension. Table 5.2, summarizes the naming-convention for RD. Table 5.3, encapsulates the RD sample-durations for signals.

FILENAME IDENTIFIERS	ASSOCIATION
Modality	Full AV – 01 Video only – 02 Audio only - 03
Vocal channel	Speech – 01 Song - 02
Emotion	Neutral – 01 Calm – 02 Happy – 03 Sad – 04 Angry – 05 Fearful – 06 Disgust – 07 Surprised - 08
Emotional intensity	Normal – 01 Strong - 02 NOTE:- There is no robust intensity for 'Neutral'
Statement	"Kids are talking by the door" – 01 "Dogs are sitting by the door" - 02
Repetition	First repetition – 01 Second repetition - 02
Actor	01 to 24 Male actors – Odd numbered Female actors – Even numbered

Table 5.2.RD naming-convention

RD	Sample duration	Length for extraction
RD - Speech	3 - 5 secs	2.5 secs
RD - Song	4 – 5 secs	3.5 secs

Table 5.3.RD durations

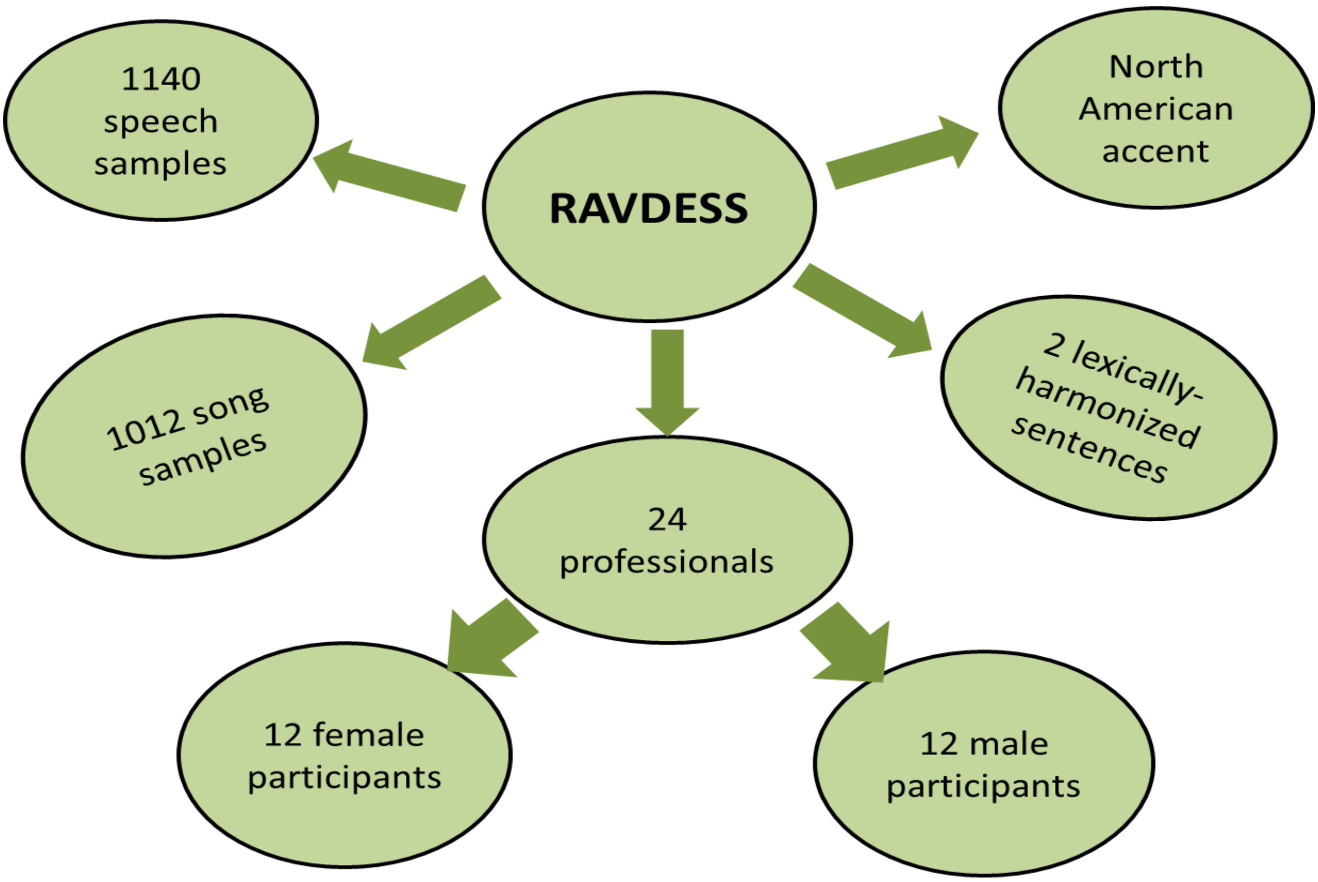


Fig. 5.1. Important features of RD

CHAPTER 6

EXPERIMENTAL RESULTS

Prior to advancing to the investigational aspect of the section, each data-instance id segregated in-to a training-set (TrS) constituting of 80-% of the data and the left-over instances are fed in-to the testing-set (TsS). Data-processing techniques like label-encoding is incorporated to sieve-out the outliers.

As mentioned beforehand, once the data is prepped feature-extraction comes into picture. Here, we aim at abstracting 182 features in total from the auditory-signals of RD. Table 6.1, portrays an example of neutral-emotion and some of the attributes that are pre-treated in order to bring all the features in a comparable and combinatorial communal form and then converted to a vector.

After the feature-extraction, the obtained attributes are sourced to an SVM classifier to get a performance yardstick. Figure 6.1, establishes the metrics obtained here.

Moving on to the HF-block of our proposal, the MEF is engineered and the attributes data is sourced to the model. Figure 6.2, shows the accuracy vs TL grid or the Pareto font on the TrS and TsS. This indicates towards the expected accuracies from HF.

Figure 6.3, gives a glimpse of the KnwB incorporated for the RD.

Figure 6.4, gives a foretaste of the member-ship function smeared on every feature of the RD. Here, only first 6 are displayed for reference.

Index	Feature	
0	0.19984	
1	-704.991	
2	65.55297	
3	-9.38088	
4	21.50494	
5	-0.27633	
6	6.958925	
7	-8.17953	
8	0.057792	
9	-12.8638	
...	...	
173	6.33E-17	
174	5.65E-17	
175	5.75E-17	
176	6.87E-17	
177	7.15E-17	
178	6.06E-17	
179	6.38E-17	
180	6.60E-17	
181	5.64E-17	
emotions	neutral	

Table 6.1.Features extracted for an example of Neutral emotion

	precision	recall	f1-score	support
0	0.88	0.82	0.85	44
1	0.70	0.74	0.72	35
2	0.72	0.76	0.74	17
accuracy			0.78	96
macro avg	0.77	0.78	0.77	96
weighted avg	0.79	0.78	0.78	96

Fig. 6.1. Results for SVM classification

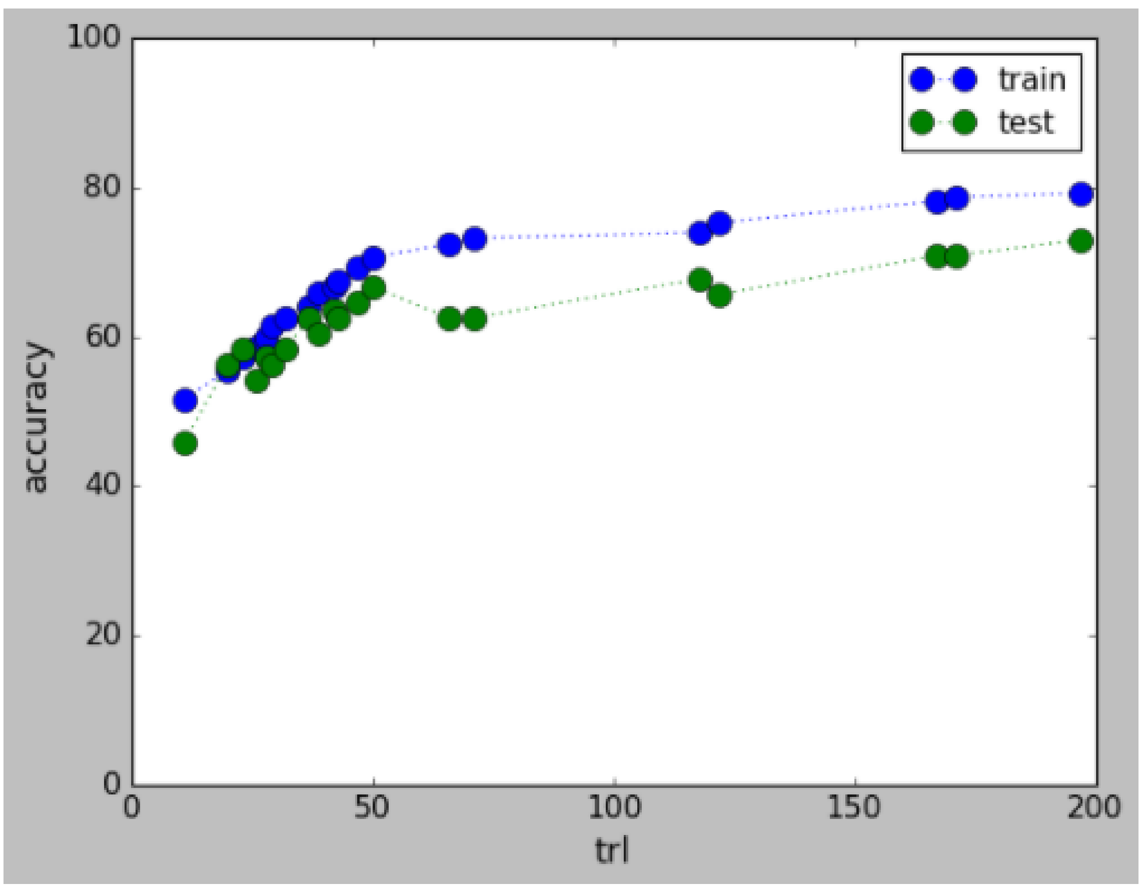


Fig. 6.2.Pareto font for test and TrS data for MEFs

```
RULE BASE
1:      IF 2 is VL  AND 6 is L  AND 20 is L THEN emotions is 2
2:      IF 2 is M  AND 4 is M  AND 6 is M  AND 8 is L THEN emot
ions is 1
3:      IF 2 is M  AND 4 is M  AND 12 is VL THEN emotions is 1
4:      IF 2 is M  AND 4 is M  AND 12 is L  AND 39 is VL THEN e
motions is 0
5:      IF 2 is M  AND 4 is M  AND 6 is M  AND 12 is M  AND 26
is L THEN emotions is 0
6:      IF 2 is M  AND 4 is M  AND 8 is H  AND 12 is H  AND 36
is VL THEN emotions is 1
7:      IF 2 is H  AND 8 is L  AND 12 is H  AND 19 is M THEN em
otions is 0
8:      IF 2 is VH  AND 19 is VL THEN emotions is 0
9:      IF 2 is H  AND 6 is L  AND 8 is M  AND 18 is M  AND 27
is H THEN emotions is 0
10:     IF 2 is H  AND 8 is H  AND 22 is L  AND 39 is M THEN em
otions is 0
11:     IF 2 is VH  AND 12 is M  AND 19 is H THEN emotions is 0
```

Fig. 6.3.KnwB or rule-base taken up

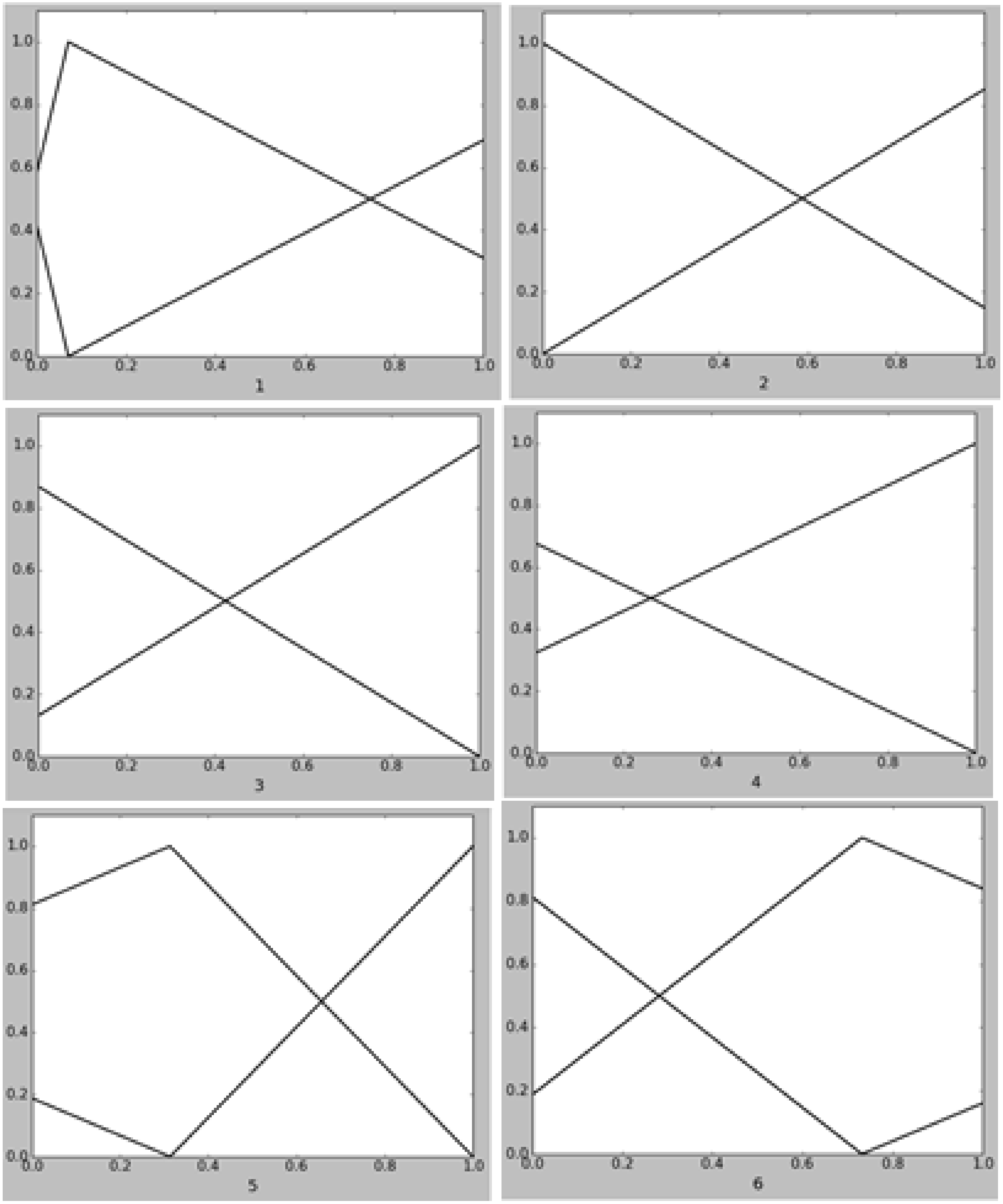
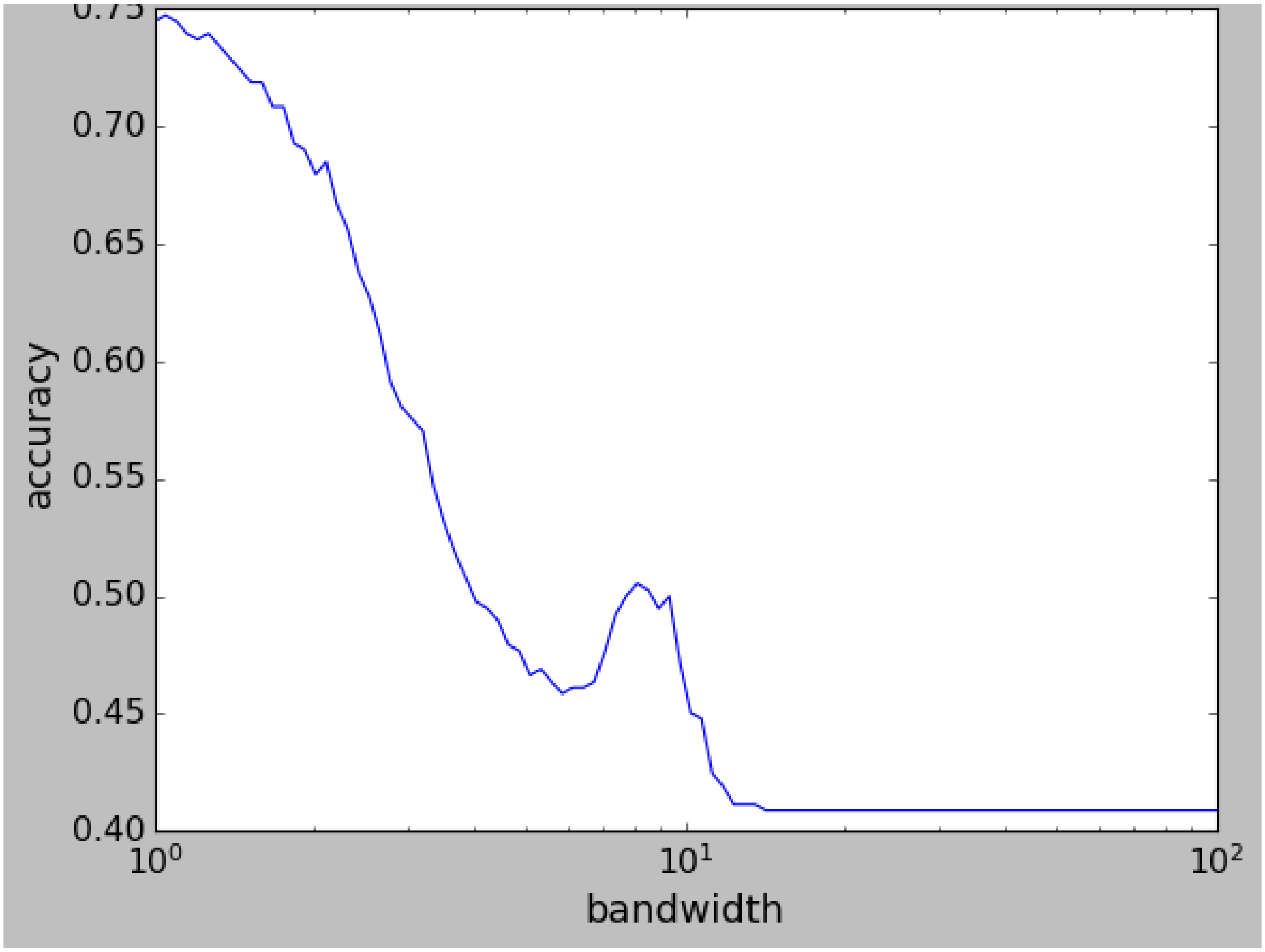


Fig. 6.4.The triangle-membership functions for first few features taken up




```
{'bandwidth': 1.0476157527896648}
accuracy = 0.7473342447026659
```

Fig. 6.5.DE performance individually

Coming to the DE-block of our proposal, which is pretty intuitive and figure 6.5, gives a foretaste of the variation of accuracy corresponding to wavering band-widths. The most optimal band-width needs to be adopted for efficient execution.

Finally, table 6.2, displays the comparison among the varied related techniques and their performance on the same data-base in terms of percentage-accuracy. This clearly states that on the given RD and among the chosen approaches, HFF-DE throws the most accurate products.

Dataset	SVM	Naïve Bayes	MEFs only	DE only	HFF-DE
RD (182 features)	78.12%	56.7%	72.9%	78.12%	79.17%

Table 6.2.Accuracies of bidding processes on RD

In the cases where the constraints of the MEFs in the HF block, are kept at low-value including but not limiting to amount of fuzzy-rules, the segregation-accuracy of HF portrays to be pretty truncated. As discussed above, this is an ordinarily observed phenomenon that the time-complexity paces-up once the expanse of fuzzy-rules is heightened. The KE, on the other hand is known to bring in high rates of accuracy thus, lifting the rates for HFF-DE scheme. This is very evidently visible that the approach in consideration, out-performs the HF individual for the same constraints being referred to.

Furthermore, the DE produces the density-functions of the given instance on the basis of an information-chauffeured tactic, rather than shadowing a manual one. This condenses the un-certainty, Moreover, abstracting the best from both and coalescing them inevitably leads to a further shrunken un-certainty over-all.

To summarize, significant advancements in the performance of the HFF-DE is evident taking the statistically driven data acquired from the results of the tryout as source.

CHAPTER 7

GROUND LAYING

FOR FUTURE WORK

The Facial expressions can be majorly thought of being categorized into three parts, Macro expressions, Micro expressions (MEx) and subtle expressions.

All the expressions cast the 7 basic emotions. Macro expressions are detected when there is no reason to modify or conceal the true emotions and last for about 0.5 to 4 seconds, on the entire face. They are easy to detect. MEx on the other end occur when the true emotions are being concealed, voluntarily or involuntarily and last for less than ½ a second. Subtle expressions are intensity of the emotion oriented and not the time for which it lasts.

When a being feels the need to masquerade his or her sincere emotions due to certain cultural display rule or doesn't feel appropriate conveying them truly [60] that is when they willfully apprehend and attempt to disguise their facial expressions. In lesser words, MEx occur when a person attempts to camouflage their true emotions [2] [59]. When they consciously realize that a facial expression is occurring, the person may try to suppress the facial expression because showing the emotion may not be appropriate or due to a cultural display rule [60]. After the suppression has occurred, the person may or may not choose to hide the original expression and display a MEx. MEx are more expected in high-stake situations as hiding true emotions become more probable.

MEx are generally short lived and thus they need to be treated specially, unlike the commonly observed Macro expressions. The typical traditional span of a MEx is nothing more than 500 ms [61]. The duration was observed to be less than 200 ms , when first discovered [62] then there were some other definitions of it lasting shorter than 250 ms [63], lesser than 330 ms [64], and shorter than ½ of a second [65]. If we consider an anatomic view then we can say that the facial muscles consist of swiftly moving fibres fibers that tighten and unwind in a span shorter than 20 ms taking in account a latency period in which they receive instructions from the nervous system [66].

Matsumoto and Hwang [67] have shown if humans could be trained to detect MEx. The ability to understand original emotions were seen to have improved but this retained only for a few weeks after a training session. Thus, training humans can be expensive and a lengthy process. Thus, a need to interact with computers emerged. The domain in which such an interaction between humans and computers is possible is known as Affective Computing. It is an interdisciplinary field that combines computer science with psychology and cognitive science. This is the technology that is based on the ability of a computer to respond to a human stimulus, preferably related to mood or emotions [68].

MEx detection and analysis of true emotions can have applications in a number of fields, one such being, security. By knowing the true emotions of a suspect can lead to efficient and much more smoother interrogations. The false statements would be much easier to detect, taking a significant leap from the classic polygraph, having evidence of deceiving with proper preparation [69]. This can also contribute in a better understanding of the nervous system and related diseases. One such disease being, Huntingston's disease, which leads to a deficiency in emotion recognition in its patients. There is not much accuracy observed in in sensing facial movements that converge at anger and disgust particularly, around a mediocre one for sadness and surprise, adding on a dearth in detecting the surprise and disgust intensities [70].

The process of programmed analysis of facial MEx can be outlined to be consisting of two groups, namely, MEx spotting and recognition methods. The MEx spotting starts with a preliminary processing round, followed by feature abstraction and then finally detection of the MEx. In introductory processing, after the facial landmarks are detected and tracked, which can be done both manually or automatically, the face registration takes place, which can be area based by using template matching or correlation or feature based using region features, line features and point features. Once the face is registered, facial regions such as, eyes, nose, mouth are masked, following which the face is split into several regions called face region retrieval.

The feature abstraction is exploited even later at the facial MEx detection stage. The subtle variations can be quantified by tracking the visually distinctive facial attributes that can occur during a MEx such as edges, corners, or other facial marks. What is more, tracing the movements from the complete facial county is a heighted computational task. We may locate some region of interests (ROIs) on the face and extracted the video cubes from each of them. Once, we have the features we can move on to detecting the MEx that includes movement spotting and apex spotting. The micro-expression recognition consists of the preliminary processing, the feature extraction, and the classification.

7.1 MEX DATASET

In order to work on any MEx detection system, a well labeled dataset is a major prerequisite. In the beginning of around the twentieth century, MEx research had picked up the pace and that is when plenty of datasets have spawned aiming at exposure and analysis of the facial MEx. Some of the earliest datasets of the emotions were non-spontaneous or posed expressions. Posed MEx do not re-create a real-world picture and hence a truncated frame rate can be assumed to be prone to risking loss of essential material. Also, the emotions represented in such datasets varied significantly from the natural emotions and thus, they were not very accurate. Then the spontaneous datasets were introduced.

A. Creating a Dataset :

A Before we dig into the existing MEx datasets, let us briefly look at the steps involved in building a dataset from scratch.

Steps involved in creating a dataset [66] are:

1) *Emotion Inducement Procedure*- along with a high stake situation. For example, to assure occurrence of MEx, participants were made to withhold their emotions by increasing the stakes, by promising a money prize to the participant who would bury his or her emotions to the most extent, and thus establishing a high-stakes situation.

2) *Equipment and Experimental Setup*- camera, lighting and high speed data capture.

3) *Inducement Stimuli*- fear inducement videos and or unexpected events.

4) *Questionnaire*- participants wait until a stimulus has been experienced and then record what emotion they felt during each stimulus.

Dataset	MEx	Emotions	Rate	Resolution	Contributors	Demographic
Polikovsky [16]	13	7	200	640x480	11	3
USF-HD [18]	100	4	29.7	720x1280	-	-
YorkDDT [19]	18	-	25	320x240	9	-
SMIC [22]	164	3	100 + 25	640x480	20	3
CASME [21]	195	7	60	640x480 + 1280 x720	35	1
CASME II [36]	247	5	200	640x480	35	1
CAS(ME)^2 [25]	250 macro+ 53 MEx	4	30	640x480	22	1
SAMM [24]	159	7	200	2040x1088	20	13

Table 7.1.1. Spontaneous and non spontaneous datasets brought together

7.2 LITERATURE REVIEW ON MEX

YEAR	LBP-TOP	HOOF	HOG-3D	DEEP LEARNING	OTHERS	TOTAL
2009-13	1	1	2	0	0	7
2014	5	0	0	0	6	11
2015	5	1	2	0	4	12
2016	4	4	1	1	6	16
2017-18	-	5	0	2	2	14

Table 7.2.1. Mex publications approximated frequency

Table 7.2.1 shows the number of publications on MEx detection. Followed by Table 7.2.2 jotting down the MEx publication and their specifications.

YEAR	AUTHORS	DATASETS	FEATURE	TOOL	OPTIMUM RESULTS
2009	Polikovsky et al. [71]	Polikovsky	HOG-3D	K-means	AUs
2011	Pfister et al. [72]	SMIC (old)	LBP-TOP	RF, MKL & SVM	MKL-71.4%
2011	Pfister et al. [73]	SPOS	CLBP-TOP	MKL, LINEAR & SVM	MKL-80%
2013	Polikovsky & Kameda [74]	Polikovsky	HOG-3D	K-means	11 AUs
2013	Li et al. [75]	SMIC (all 3)	LBP-TOP	SVM	VIS-52.11%
2013	Song et al. [76]	SEMAINE	HOF & HOG	SVR	-
2014	Guo et al. [77]	SMIC	LBP-TOP	Nearest neighbor	65.83%
2014	Yan et al. [78]	CASME II	LBP-TOP	SVM	63.41%
2014	Wang et al. [79]	CASME & CASME II	TICS	SVM	CASME-61.85% CASME II-58.53%
2014	Le et al. [80]	CASME II & SMIC	LPB-TOP & STM	AdaBoost	SMIC-44.34% CASME II-43.78%

Table 7.2.2. Mex publications 2009-2020 (1)

2014	Lu et al. [81]	CASME B, CAMSE II & SMIC	DTCM	RF & SVM	CASME-64.95% CASME II-64.19% SMIC-82.86%
2014	Liong et al. [82]	SMIC & CASME II	OSW-LBP-TOP	SVM	SMIC-57.54% CASME II-66.40%
2014	Wang et al. [83]	CASME	DTSA	ELM	46.90%
2014	Davison et al. [84]	CAMSE II	LBP-TOP & GDs	SVM & RF	RF-92.6%
2015	House & Meyer [85]	SMIC	LGCP-TOP	SVM	48.10%
2015	Wang et al. [86]	SMIC & CASME II	LBP-SIP & LBP-MOP	SVM	CASME & LBP-MOP-66.8%
2015	Wang et al. [87]	CASME & CASME II	TICS, CIELab & CIELuv	SVM	CASME-61.86% CASME II-62.30%
2015	Le et al. [88]	CASME II	DMDSP, LBP-TOP	SVM & LDA	0.52 F1 Score
2015	Huang et al. [89]	SMIC & CASME II	STLBP-IP	SVM	SMIC-57.93% CASME II-59.51%
2015	Li et al. [90]	SMIC & CASME II	HIGO, HOG & LBP	LSVM	SMIC-53.52% CASME II-57.49%
2015	Kamarol et al. [91]	CASME II	STTM	SVM 1-on-1	91.71%
2016	Liu et al. [92]	SMIC, CASME & CASME II	MDMO	SVM	CASME-68.86% CASME II-67.37% SMIC- 80%
2016	Chen et al. [93]	CASME II (36)	HOG-3D	Fuzzy	86.67%
2016	Talukder et al. [94]	SMIC	LBP-TOP	SVM	SMIC-NIR-62%
2016	Duan et al. [95]	CASME II	LPB-TOP (eye)	26 Classifiers	-

Table 7.2.2. Mex publications 2009-2020 (2)

YEAR	AUTHORS	DATASETS	FEATURE	TOOL	OPTIMUM RESULTS
2016	Huang et al. [96]	SMIC, CASME & CASME II	STLBP-IP- new	SVM	CASME-64.33% SMIC-63.41% CASME II-64.78%
2016	Wang et al. [97]	CASME II	LBP-TOP	KNN & SVM	75.30%
2016	Zhang et al. [98]	CASME II	LDA & PCA (gabor filter)	SVM	-
2016	Huang et al. [99]	SMIC, CASME & CASME II	STCLQP	Codebook	CASME-57.31% SMIC-64.02% CASME II-58.39%
2016	Ben et al. [100]	CASME	MMPTR	Euclidean Dist.	80.2%
2016	Liong et al. [101]	SMIC & CASME II	Bi-WOOF	SVM	(F1 Score) SMIC-HS-0.62 CASME II-0.61
2016	Liong et al. [102]	SMIC & CASME II	Bi-WOOF	SVM	F1-Score- CASME II- 0.59 Accuracy- SMIC-VIS-53.52%
2016	Liiong et al. [103]	SMIC & CASME II	Optical Strain	SVM	SMIC- 52.44% CASME II-63.41%
2016	Oh et al. [104]	SMIC & CASME II	I2D	SVM	(F1 Score) SMIC- 0.44 CASME II-0.41

Table 7.2.2. Mex publications 2009-2020 (3)

2016	Wang et al. [105]	CASME & CASME II	STCCA	SVM & Nearest neighbor	(Mean Recognition Accuracy) CASME- 41.20% CASME II- 38.39%
2016	Zheng et al. [106]	CASME & CASME II	HOOF & LBP-TOP	RK-SVD	CASME-69.04% CASME II-63.25%
2016	Kim et al. [107]	CASME II	CNN	LSTM	60.98%
2016	Mayya et al. [108]	SMIC & CASME II	TIM & DCNN	SVM	CASME II- 67.742% SMIC- 78.05%
2017	Zhang et al. [109]	CASME II	Optical Flow & LBP-TOP	RF, KNN & SVM	62.50%
2017	Zheng [110]	SMIC, CASME & CASME II	GSR-2D	SRC	CASME-71.19% CASME II-64.88%
2017	Ben et al. [111]	CASME II	HWP-TOP	SSVM	Recognition Rate- 0.868
2017	Zong et al. [112]	SMIC & CASME II	LBP-TOP	TSRG	UAR- 60.15
2017	Happy & Routray [113]	CASME II, SMIC & CASME	FHOFO	LDA, KNN & SVM	(F1 Score) CASME-05489 SMIC- 0.5243 CASME II- 0.5248
2017	Hao et al. [114]	JAFFE	DBN & WLD	DBN	Recognition Rate- 92.66

Table 7.2.2. Mex publications 2009-2020 (4)

2017	Peng et al. [115]	CASME / CASME II	OF	DTS-CNN	66.67%
2017	Feng Xu et al. [116]	SMIC, CASME & CASME II	FDM, LBP-TOP & DTSA	SVM	FDM-CASME-42.02% CASME II- 41.96% SMIC- 75.66%
2018	Liong et al. [117]	SMIC & CASME II	LBP-TOP & OSF	SVM	(F-measure) SMIC- 0.51 CASME II-0.31
2018	Zhu et al. [118]	CASME II	OF & LBP-TOP	SVM	53.3%
2018	Zong et al. [119]	CASME II & SMIC	LBP Variants	KGSL	F1- CASME II- 0.6125
2018	Gan et al. [120]	CASME II & SMIC	BiVACNN	CNN	Recognition Accuracy- 80%
2018	Dong et al. [121]	CK+	2D landmark feature	LSTM	77%
2019	Kai et al. [122]	CASME II	3 Facial landmarks	Euclidean Dist.	82.30%
2019	Huang et al. [123]	SMIC, CASME & CASME II	STLBP-RIP & DiSTLBP-RIP	-	62.75 % & 64.78% respectively
2020	Zong et al. [124]	SMIC & CASME II	ASSM & TTRM	SVM	SMIC- 76.29% CASME II- 55.71%

Table 7.2.2. Mex publications 2009-2020 (5)

7.3 CHALLENGES AROUND MEX

As evident from the previous sections, considerable amount of research has been done in the discipline of MEx. Regardless, there exist few challenges that can be considered for upcoming works.

The introduction to this paper highlights the biggest encounter that one could face when handling MEx in general. MEx are extremely transient in nature as they live for maximum about 0.5s. Moreover, not just that they exist for a small duration, they are also of truncated intensity. Thus, making them indistinguishable through naked eyes and sometimes even by the computing algorithms. One way out in this case is generally by adding a pre-processing stage in order to intensify the MEx. Recorded

There are some publicly accessible datasets that aim at keeping MEx under the spotlight. However, there has also been observed some prejudice in them as even in the spontaneous datasets, the MEx are outcomes of forced feelings from an external source and hence, vary from the natural or actual MEx numerously. In addition, a meticulous milieu with severe circumstances is chosen, in general to get the required information. Thus, even the well-tested algorithms might not produce the appropriate conclusions under typical settings.

Head-pose deviations and radiance variations are indeed the most common and intricate occurrence along with being unavoidable in the process of capturing reactions for datasets. These fall into the environmental differences and cannot be assumed to be negligible as some features depend on the intensity of pixels and with illumination changes the process may lead to erroneous valuations. Uncalled for crown maneuvers may come closely to an existing class during classification and be misjudged as some MEx. This could also hinder the accuracy factor.

CHAPTER 8

CONCLUSION & FUTURE WORK

The HFF-DE is a renewed method that has been distended as an enhancement scheme by supporting the conception of a fusion of a hybrid-fuzzy concept which is supported on fuzzy-rule-based classification-system alongside with density-estimation technique that in-turn incorporates a robust station of naïve Bayes classifier as a structural module. The fusion is launched through the support of a scheme of voting-classifier to highlight the decision formulating act. This implementation has been done particularly in the field of Auditory Affect Recognition. The DE aims at enhancing the flexibility of the system whereas; the HF caters to the issue of uncertainty. Therefore, utilizing the framework of voting classifier, the 2 techniques have been meticulously combined by maintaining the spot light on their corresponding pluses to broaden the improved accuracy and sturdiness.

Moreover, the perception of sentiments bridged through speech is really a moderately latest exploration matter within turf of voice administering while it is inspected throughout lone couple of preceding times. Certainly, aforementioned has acknowledged decent sum of attentiveness, including but not limited to academia, industry insides, owing to the amplified enactment along with dependability over-all.

This is our belief that this work would only form a robust base for many great opportunities opening in the field of auditory affect detection. To extend this work to greater heights, the HFF-DE model can be applied on other datasets and a comparison can be drawn to depict the precision and reliability of the approach. Another huge milestone would be to outspread HFF-DE and reach the visual domain of the affect recognition domain of research. Hence, truly elaborating the rightful potential of HFF-DE.

REFERENCES

[1] P. Ekman,: An argument for basic emotions. Cognition and Emotion, vol. 6, pp. 169–200, 1992.

[2] Paul Ekman, Emotions Revealed: Understanding Faces and Feelings. Phoenix, 2004.

[3] P. Ekman and E. L. Rosenberg, What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS), ser. Series in Affective Science. Oxford University Press, 2005.

[4] Pedrycz, W., Wang, X.: Designing fuzzy sets with the use of the parametric principle of justifiable granularity. IEEE Trans. Fuzzy Syst. 24(2), 489–496 (2016).

[5] Liao, H., Mi, X., Xu, Z.: A survey of decision-making methods with probabilistic linguistic information: Bibliometrics, preliminaries, methodologies, applications and future directions. Fuzzy Optimiz. Decis. Mak. (2019).

[6] Ferna´ndez, A., Altalhi, A., Alshomrani, S., Herrera, F.: Why linguistic fuzzy rule based classification systems perform well in big data applications? Int. J. Comput. Intell. Syst. 10(1), 1211–1225 (2017).

[7] Pelusi, D., Mascella, R., Tallini, L., Nayak, J., Naik, B., Abraham, A.: Neural network and fuzzy system for the tuning of gravitational search algorithm parameters. Exp. Syst. Appl. 102, 234–244 (2018).

[8] Jiang, W., Zhang, Z., Deng, X.: A novel failure mode and effects analysis method based on fuzzy evidential reasoning rules. IEEE Access 7, 113605 113615 (2019).

[9] Morente-Molinera, J., Wu, X., Morfeq, A., Al-Hmouz, R., Herrera- Viedma, E.: A novel multi-criteria group decision-making method for heterogeneous and dynamic contexts using multigranular fuzzy linguistic modelling and consensus measures. Inform. Fusion 53, 240–250 (2020).

[10] Liu, P., Zhang, X.: A new hesitant fuzzy linguistic approach for multiple attribute decision making based on Dempster-Shafer evidence theory. Appl. Soft Comput. 86, 105897 (2020).

[11] Xingjian Song, Bowen Qin, Fuyuan Xiao1: FR–KDE:A Hybrid Fuzzy Rule-Based Information Fusion Method with its Application in Biomedical Classification. Int. J. Fuzzy Syst. https://doi.org/10.1007/s40815-020-00957-z (2020).

[12] Xiao, F., Zhang, Z., Abawajy, J.: Workflow scheduling in distributed systems under fuzzy environment. J. Intell. Fuzzy Syst. 37(4), 5323–5333 (2019).

[13] Xiao, F.: EFMCDM: Evidential fuzzy multi criteria decision making based on belief entropy. IEEE Trans. Fuzzy Syst. 28(7), 1477–1491 (2020).

[14] Fei, L., Feng, Y., Liu, L.: On Pythagorean fuzzy decision making using soft likelihood functions. Int. J. Intell. Syst. 34(12), 3317–3335 (2019).

[15] Cl, Fan, Song, Y., Fu, Q., Lei, L., Wang, X.: New operators for aggregating intuitionistic fuzzy information with their application in decision making. IEEE Access 6, 27214–27238 (2018).

[16] Fei, L., Deng, Y.: Multi-criteria decision making in Pythagorean fuzzy environment. Appl. Intell. 50(2), 537–561 (2020).

[17] Garg, H., Rani, D.: Novel exponential divergence measure of complex intuitionistic fuzzy sets with an application to decision making process. Scientia Iranica (2019). https://doi.org/10.24200/ sci.2019.53194.3103.

[18] Liao, H., Wu, X.: DNMA: A double normalization-based multiple aggregation method for multi-expert multi-criteria decision making. Omega (2019). https://doi.org/10.1016/j.omega.2019.04. 001.

[19] Qin, B., Xiao, F.: A non-parametric method to determine basic probability assignment based on Kernel Density Estimation. IEEE Access 6, 73509–73519 (2018).

[20] Meng, D., Liu, M., Yang, S., Zhang, H., Ding, R.: A fluid structure analysis approach and its application in the uncertainty based multidisciplinary design and optimization for blades. Adv. Mech. Eng. 10(6), 1687814018783410 (2018).

[21] Cao, Z., Chuang, C.H., King, J.K., Lin, C.T.: Multi-channel EEG recordings during a sustained-attention driving task. Scientific Data (2019). https://doi.org/10.1038/s41597-019-0027-4.

[22] Fei, L., Lu, J., Feng, Y.: An extended best-worst multi-criteria decision-making method by belief functions and its applications in hospital service evaluation. Comput. Industr. Eng. 142, 106355 (2020).

[23] Garg, H., Rani, D.: Multi-criteria decision making method based on bonferroni mean aggregation operators of complex intuitionistic fuzzy numbers. J. Industr. Manag. Optimiz. (2020). https://doi.org/10.3934/jimo.2020069.

[24] Pan, L., Deng, Y.: An association coefficient of belief function and its application in target recognition system. Int. J. Intell. Syst. 35, 85–104 (2020).

[25] Xu, X., Li, S., Song, X., Wen, C., Xu, D.: The optimal design of industrial alarm systems based on evidence theory. Control Eng. Pract. 46, 142–156 (2016).

[26] Li, Y.F., Huang, H.Z., Mi, J., Peng, W., Han, X.: Reliability analysis of multi-state systems with common cause failures based on bayesian network and fuzzy probability. Ann. Oper. Res.(2019).

[27] Alcantud, J.C., Feng, F., Yager, R.: An N-soft set approach to rough sets. IEEE Trans. Fuzzy Syst. (2019). https://doi.org/10.1109/TFUZZ.2019.2946526.

[28] Deng, Y.: Uncertainty measure in evidence theory. Sci. China Inform. Sci. (2021). https://doi.org/10.1007/s11432-020-3006-9.

[29] Zhang, H., Deng, Y.: Weighted belief function of sensor data fusion in engine fault diagnosis. Soft Comput. 24(3), 2329–2339 (2020).

[30] Cao, X., Deng, Y.: A new geometric mean FMEA method based on information quality. IEEE Access 7(1), 95547–95554 (2019).

[31] Xu, X., Li, S., Song, X., Wen, C., Xu, D.: The optimal design of industrial alarm systems based on evidence theory. Control Eng. Pract. 46, 142–156 (2016).

[32] Xu, X., Xu, H., Wen, C., Li, J., Hou, P., Zhang, J.: A belief rule based evidence updating method for industrial alarm system design. Control Eng. Pract. 81, 73–84 (2018).

[33] Meng, D., Liu, M., Yang, S., Zhang, H., Ding, R.: A fluid structure analysis approach and its application in the uncertaintybased multidisciplinary design and optimization for blades. Adv. Mech. Eng. 10(6), 1687814018783410 (2018).

[34] Wu, X., Liao, H.: A consensus-based probabilistic linguistic gained and lost dominance score method. Eur. J. Oper. Res. 272(3), 1017–1027 (2019).

[35] Fang, R., Liao, H., Yang, J.B., Xu, D.L.: Generalised probabilistic linguistic evidential reasoning approach for multi-criteria decision-making under uncertainty. J. Oper. Res. Soc. (2019). https://doi.org/10.1080/01605682.2019.1654415.

[36] Gallo G., Ferrari V., Marcelloni F., Ducange P. (2020) SK-MOEFS: A Library in Python for Designing Accurate and Explainable Fuzzy Models. In: Lesot MJ. et al. (eds) Information Processing and Management of Uncertainty in Knowledge-Based Systems. IPMU 2020. Communications in Computer and Information Science, vol 1239. Springer, Cham. https://doi.org/10.1007/978-3-030-50153-2_6.

[37] Antonelli, M., Ducange, P., Marcelloni, F.: Multi-objective evolutionary design of fuzzy rule-based systems. In: Handbook on Computational Intelligence: Volume 2: Evolutionary Computation, Hybrid Systems, and Applications, pp. 635–670. World Scientific (2016).

[38] Fazzolari, M., Alcala, R., Nojima, Y., Ishibuchi, H., Herrera, F.: A review of the application of multiobjective evolutionary fuzzy systems: current status and further directions. IEEE Trans. Fuzzy Syst. 21(1), 45–65 (2012).

[39] Coello, C.A.C., Lamont, G.B.: Applications of Multi-objective Evolutionary Algorithms, vol. 1. World Scientific, Singapore (2004).

[40] Ducange, P., Marcelloni, F.: Multi-objective evolutionary fuzzy systems. In: Fanelli, A.M., Pedrycz, W., Petrosino, A. (eds.) WILF 2011. LNCS (LNAI), vol. 6857, pp. 83–90. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23713-3_11.

[41] Qin, B., Xiao, F.: A non-parametric method to determine basic probability assignment based on Kernel Density Estimation. IEEE Access 6, 73509–73519 (2018).

[42] Carletti, M., Masiero, C., Beghi, A., Susto, G.A.: Explainable machine learning in industry 4.0: evaluating feature importance in anomaly detection to enable root cause analysis. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), pp. 21–26. IEEE (2019).

[43] Fernandez, A., Herrera, F., Cordon, O., del Jesus, M.J., Marcelloni, F.: Evolutionary fuzzy systems for explainable artificial intelligence: why, when, what for, and where to? IEEE Comput. Intell. Mag. **14**(1), 69–81 (2019).

[44] Pancho, D.P., Alonso, J.M., Magdalena, L.: Quest for interpretability-accuracy trade-off supported by fingrams into the fuzzy modeling tool GUAJE. Int. J. Comput. Intell. Syst. **6**(sup1), 46–60 (2013).

[45] Alonso, J.M., Bugarín, A.: ExpliClas: automatic generation of explanations in natural language for weka classifiers. In: 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–6. IEEE (2019).

[46] Antonelli, M., Ducange, P., Marcelloni, F.: A fast and efficient multi-objective evolutionary learning scheme for fuzzy rule-based classifiers. Inf. Sci. **283**, 36–54 (2014).

[47] Ducange, P., Mannarà, G., Marcelloni, F.: Multi-objective evolutionary granular rule-based classifiers: an experimental comparison. In: 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–6. IEEE (2017).

[48] Knowles JD, Corne DW Approximating the nondominated front using the Pareto archived evolution strategy. Evol Comput 8(2):149–172 (2000).

[49] Antonelli M, Ducange P, Lazzerini B, Marcelloni F : Learning concurrently data and rule bases of Mamdani fuzzy rule-based systems by exploiting a novel interpretability index. Soft Comput 15(10):1981–1998 (2011).

[50] Antonelli, M., Ducange, P., Lazzerini, B. *et al.* : Multi-objective evolutionary design of granular rule-based classifiers. *Granul. Comput.* **1,** 37–58 . https://doi.org/10.1007/s41066-015-0004-z. (2016).

[51] Antonelli M, Ducange P, Lazzerini B, Marcelloni F : Learning knowledge bases of multi-objective evolutionary fuzzy systems by simultaneously optimizing accuracy, complexity and partition integrity. Soft Comput 15(12):2335–2354 (2011).

[52] Antonelli M, Ducange P, Marcelloni F : Genetic training instance selection in multi-objective evolutionary fuzzy systems: a co-evolutionary approach. IEEE Trans Fuzzy Syst 20(2):276–290 (2012).

[53] Antonelli M, Ducange P, Marcelloni F : An experimental study on evolutionary fuzzy classifiers designed for managing imbalanced datasets. Neuro computing 146:125–136 (2014).

[54] Liu, Z., Liu, Y., Dezert, J., Cuzzolin, F.: Evidence combination based on credal belief redistribution for pattern classification. IEEE Trans. Fuzzy Syst. (2019). https://doi.org/10.1109/TFUZZ.2019.2911915.

[55] Xiao, F.: A distance measure for intuitionistic fuzzy sets and its application to pattern classification problems. IEEE Trans. Syst. Man Cybern. (2019). https://doi.org/10.1109/TSMC.2019.2958635.

[56] Xu, X., Zheng, J., Yang Jb, Xu, Dl, Chen Yw: Data classification using evidence reasoning rule. Knowledge Based Syst. 116, 144–151 (2017).

[57] Jiang, W., Huang, K., Geng, J., Deng, X.: Multi-scale metric learning for few-shot learning. IEEE Trans Circ. Syst. Video Technol. (2020). https://doi.org/10.1109/TCSVT.2020.2995754.

[58] Steven R. Livingstone1,2*, Frank A. Russo1: 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 | https://doi.org/10.1371/journal.pone.0196391 (2018).

[59] P. Ekman, “Lie catching and microexpressions,” in The Philosophy of Deception, C. W. Martin, Ed. New York, NY, USA: Oxford Univ. Press, 2009, pp. 118–133.

[60] D. Matsumoto, S. H. Yoo, and S. Nakagawa, “Culture, emotion regulation, and adjustment.” J. Pers. Soc. Psychol., vol. 94, no. 6, pp. 925–937, 2008.

[61] W.-J. Yan, Q. Wu, J. Liang, Y.-H. Chen, and X. Fu, “How fast are the leaked facial expressions: The duration of micro-expressions,” J. Nonverbal Behav., vol. 37, no. 4, pp. 217–230, 2013.

[62] P. Ekman and W. V. Friesen, “Nonverbal leakage and clues to deception,” Psychiatry, vol. 32, no. 1, pp. 88–106, 1969.

[63] P. Ekman, Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage. New York, NY, USA: Norton, 2001

[64] P. Ekman and E. L. Rosenberg, What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System. New York,NY, USA: Oxford Univ. Press, 2005.

[65] M. G. Frank, C. J. Maccario, and V. l. Govindaraju, “Behavior and security,” in Protecting Airline Passengers in the Age of Terrorism. Goleta, CA, USA: Greenwood Pub. Group, 2009.

[66] Adrian K. Davison, Cliff Lansley, Nicholas Costen, Kevin Tan, and Moi Hoon Yap, “SAMM: A Spontaneous Micro-Facial Movement Dataset”, IEEE Transactions on Affective Computing, vol 9, No. 1, January-March 2018.

[67] D. Matsumoto and H. S. Hwang, “Evidence for training the ability to read microexpressions of emotion,” Motivation Emotion, vol. 35, pp. 181–191, 2011.

[68] Loredana Stanciu1, Adriana Albu, “Analysis on Emotion Detection and Recognition Methods using Facial Microexpressions. A Review”, The 7th IEEE International Conference on E-Health and Bioengineering - EHB 2019.

[69] S.-J. Wang, W.-J. Yan, X. Li, G. Zhao, C.-G. Zhou et al., “Microexpression recognition using color spaces”, IEEE Transactions on Image Processing, 24(12), 2015, pp. 6034–6047.

[70] D. Matsumoto, J. LeRoux, C. Wilson-Cohn et al., “A new test to measure emotion recognition ability: Matsumoto and Ekman’s Japanese and Caucasian Brief Affect Recognition Test (JACBART)”, Journal of Nonverbal Behavior, 24(3), 2000, pp. 179-20.

[71] S. Polikovsky, Y. Kameda, and Y. Ohta, “Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor,” in Crime Detection and Prevention (ICDP 2009), 3rd International Conference on. IET, 2009, pp. 1–6.

[72] T. Pfister, X. Li, G. Zhao, and M. Pietikainen, “Recognising spontaneous facial micro-expressions,” in Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011, pp. 1449–1456.

[73] T. Pfister, X. Li, G. Zhao, and M. Pietikainen, “Differentiating ¨ spontaneous from posed facial expressions within a generic facial expression recognition framework,” in Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on. IEEE, 2011, pp. 868–875.

[74] S. Polikovsky and Y. Kameda, “Facial micro-expression detection in hi-speed video based on facial action coding system (facs),” IEICE transactions on information and systems, vol. 96, no. 1, pp. 81–92, 2013.

[75] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikainen, “A spontaneous micro-expression database: Inducement, collection and baseline,” in Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on. IEEE, 2013, pp. 1–6.

[76] Y. Song, L.-P. Morency, and R. Davis, “Learning a sparse codebook of facial and body microexpressions for emotion recognition,” in Proceedings of the 15th ACM on International conference on multimodal interaction. ACM, 2013, pp. 237–244.

[77] Y. Guo, Y. Tian, X. Gao, and X. Zhang, “Micro-expression recognition based on local binary patterns from three orthogonal planes and nearest neighbor method,” in Neural Networks (IJCNN), 2014 International Joint Conference on. IEEE, 2014, pp. 3473–3479.

[78] W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, and X. Fu, “Casme ii: An improved spontaneous micro-expression database and the baseline evaluation,” PloS one, vol. 9, no. 1, 2014.

[79] S.-J. Wang, W.-J. Yan, X. Li, G. Zhao, and X. Fu, “Micro-expression recognition using dynamic textures on tensor independent color space,” in Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE, 2014, pp. 4678–4683.

[80] A. C. Le Ngo, R. C.-W. Phan, and J. See, “Spontaneous subtle expression recognition: Imbalanced databases and solutions,” in Computer Vision–ACCV 2014. Springer, 2014, pp. 33–48.

[81] Z. Lu, Z. Luo, H. Zheng, J. Chen, and W. Li, “A delaunaybased temporal coding model for micro-expression recognition,” in Asian Conference on Computer Vision. Springer, 2014, pp. 698– 711.

[82] S.-T. Liong, J. See, R. C.-W. Phan, A. C. Le Ngo, Y.-H. Oh, and K. Wong, “Subtle expression recognition using optical strain weighted features,” in Asian Conference on Computer Vision. Springer, 2014, pp. 644–657.

[83] S.-J. Wang, H.-L. Chen, W.-J. Yan, Y.-H. Chen, and X. Fu, “Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine,” Neural processing letters, vol. 39, no. 1, pp. 25–43, 2014.

[84] A. K. Davison, M. H. Yap, N. Costen, K. Tan, C. Lansley, and D. Leightley, “Micro-facial movements: An investigation on spatio-temporal descriptors,” in Computer Vision-ECCV 2014 Workshops. Springer, 2014, pp. 111–123.

[85] C. House and R. Meyer, “Preprocessing and descriptor features for facial micro-expression recognition,” 2015.

[86] Y. Wang, J. See, R. C.-W. Phan, and Y.-H. Oh, “Efficient spatiotemporal local binary patterns for spontaneous facial microexpression recognition,” PloS one, vol. 10, no. 5, p. e0124674, 2015.

[87] S.-J. Wang, W.-J. Yan, X. Li, G. Zhao, C.-G. Zhou, X. Fu, M. Yang, and J. Tao, “Micro-expression recognition using color spaces,” IEEE Transactions on Image Processing, vol. 24, no. 12, pp. 6034– 6047, 2015.

[88] A. C. Le Ngo, S.-T. Liong, J. See, and R. C.-W. Phan, “Are subtle expressions too sparse to recognize?” in 2015 IEEE International Conference on Digital Signal Processing (DSP). IEEE, 2015, pp. 1246–1250.

[89] X. Huang, S.-J. Wang, G. Zhao, and M. Piteikainen, “Facial microexpression recognition using spatiotemporal local binary pattern with integral projection,” in Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015, pp. 1–9.

[90] X. Li, X. Hong, A. Moilanen, X. Huang, T. Pfister, G. Zhao, and M. Pietikainen, “Reading hidden emotions: spontaneous ¨ micro-expression spotting and recognition,” arXiv preprint arXiv:1511.00423, 2015.

[91] S. K. A. Kamarol, N. S. Meli, M. H. Jaward, and N. Kamrani, “Spatio-temporal texture-based feature extraction for spontaneous facial expression recognition,” in Machine Vision Applications (MVA), 2015 14th IAPR International Conference on. IEEE, 2015, pp. 467–470.

[92] Y.-J. Liu, J.-K. Zhang, W.-J. Yan, S.-J. Wang, G. Zhao, and X. Fu, "A main directional mean optical flow feature for spontaneous micro-expression recognition," IEEE Transaction on Affective Computing, 2016.

[93] M. Chen, H. T. Ma, J. Li, and H. Wang, "Emotion recognition using fixed length micro-expressions sequence and weighting method," in Real-time Computing and Robotics (RCAR), IEEE International Conference on. IEEE, 2016, pp. 427–430.

[94] B. B. Talukder, B. Chowdhury, T. Howlader, and S. M. Rahman, "Intelligent recognition of spontaneous expression using motion magnification of spatio-temporal data," in Pacific-Asia Workshop on Intelligence and Security Informatics. Springer, 2016, pp. 114– 128.

[95] X. Duan, Q. Dai, X. Wang, Y. Wang, and Z. Hua, "Recognizing spontaneous micro-expression from eye region," Neurocomputing, vol. 217, pp. 27–36, 2016.

[96] X. Huang, S. Wang, X. Liu, G. Zhao, X. Feng, and M. Pietikainen, "Spontaneous facial micro-expression recognition using discriminative spatiotemporal local binary pattern with an improved integral projection," arXiv preprint arXiv:1608.02255, 2016.

[97] Y. Wang, J. See, Y.-H. Oh, R. C.-W. Phan, Y. Rahulamathavan, H.-C. Ling, S.-W. Tan, and X. Li, "Effective recognition of facial micro-expressions with video motion magnification," Multimedia Tools and Applications, pp. 1–26, 2016.

[98] P. Zhang, X. Ben, R. Yan, C. Wu, and C. Guo, "Micro-expression recognition system," Optik-International Journal for Light and Electron Optics, vol. 127, no. 3, pp. 1395–1400, 2016.

[99] X. Huang, G. Zhao, X. Hong, W. Zheng, and M. Pietikainen, ¨ "Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns," Neurocomputing, vol. 175, pp. 564–578, 2016.

[100] X. Ben, P. Zhang, R. Yan, M. Yang, and G. Ge, "Gait recognition and micro-expression recognition based on maximum margin projection with tensor representation," Neural Computing and Applications, vol. 27, no. 8, pp. 2629–2646, 2016.

[101] S.-T. Liong, J. See, R. C.-W. Phan, and K. Wong, "Less is more: Micro-expression recognition from video using apex frame," arXiv preprint arXiv:1606.01721, 2016.

[102] S.-T. Liong, J. See, K. Wong, and R. C.-W. Phan, "Automatic micro-expression recognition from long video using a single spotted apex."

[103] S.-T. Liong, J. See, R. C.-W. Phan, Y.-H. Oh, A. C. Le Ngo, K. Wong, and S.-W. Tan, "Spontaneous subtle expression detection and recognition based on facial strain," Signal Processing: Image Communication, vol. 47, pp. 170–182, 2016.

[104] Y.-H. Oh, A. C. Le Ngo, R. C.-W. Phari, J. See, and H.-C. Ling, “Intrinsic two-dimensional local structures for micro-expression recognition,” in Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on. IEEE, 2016, pp. 1851–1855.

[105] S.-J. Wang, W.-J. Yan, T. Sun, G. Zhao, and X. Fu, “Sparse tensor canonical correlation analysis for micro-expression recognition,” Neurocomputing, vol. 214, pp. 218–232, 2016.

[106] H. Zheng, X. Geng, and Z. Yang, “A relaxed k-svd algorithm for spontaneous micro-expression recognition,” in Pacific Rim International Conference on Artificial Intelligence. Springer, 2016, pp. 692–699.

[107] D. H. Kim, W. J. Baddar, and Y. M. Ro, “Micro-expression recognition with expression-state constrained spatio-temporal feature representations,” in Proceedings of the 2016 ACM on Multimedia Conference. ACM, 2016, pp. 382–386.

[108] Veena Mayya, Radhika M. Pai and Manohara Pai M.M., “Combining Temporal Interpolation and DCNN For Faster Recognition of Micro-expressions in Video Sequences,” in Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, Jaipur, India.

[109] S. Zhang, B. Feng, Z. Chen, and X. Huang, “Micro-expression recognition by aggregating local spatio-temporal patterns,” in International Conference on Multimedia Modeling. Springer, 2017, pp. 638–648.

[110] H. Zheng, “Micro-expression recognition based on 2d gabor filter and sparse representation,” in Journal of Physics: Conference Series, vol. 787, no. 1. IOP Publishing, 2017, p. 012013.

[111] X. Ben, X. Jia, R. Yan, X. Zhang, and W. Meng, “Learning effective binary descriptors for micro-expression recognition transferred by macro-information,” Pattern Recognition Letters, 2017.

[112] Y. Zong, X. Huang, W. Zheng, Z. Cui, and G. Zhao, “Learning a target sample re-generator for cross-database micro-expression recognition,” in Proceedings of the 2017 ACM on Multimedia Conference. ACM, 2017, pp. 872–880.

[113] S. Happy and A. Routray, “Fuzzy histogram of optical flow orientations for micro-expression recognition,” IEEE Transactions on Affective Computing, 2017.

[114] X.-l. Hao and M. Tian, “Deep belief network based on double weber local descriptor in micro-expression recognition,” in Advanced Multimedia and Ubiquitous Engineering. Springer, 2017, pp. 419–425.

[115] M. Peng, C. Wang, T. Chen, G. Liu, and X. Fu, “Dual temporal scale convolutional neural network for micro-expression recognition,” Frontiers in psychology, vol. 8, p. 1745, 2017.

[116] Feng Xu, Junping Zhang, Member, Ieee, And James Z. Wang,” Microexpression Identification And Categorization Using A Facial Dynamics Map,” In Ieee Transactions On Affective Computing, Vol. 8, No. 2, April-June 2017.

[117] S.-T. Liong, J. See, R. C.-W. Phan, K. Wong, and S.-W. Tan, “Hybrid facial regions extraction for micro-expression recognition system,” Journal of Signal Processing Systems, vol. 90, no. 4, pp. 601–617, 2018.

[118] X. Zhu, X. Ben, S. Liu, R. Yan, and W. Meng, “Coupled source domain targetized with updating tag vectors for micro-expression recognition,” Multimedia Tools and Applications, vol. 77, no. 3, pp. 3105–3124, 2018.

[119] Y. Zong, X. Huang, W. Zheng, Z. Cui, and G. Zhao, “Learning from hierarchical spatiotemporal descriptors for microexpression recognition,” IEEE Transactions on Multimedia, 2018.

[120] Y.S. Gan and Sze-Teng Liong, “Bi-directional Vectors from Apex in CNN for Micro-expression Recognition,” in 2018 3rd IEEE International Conference on Image, Vision and Computing.

[121] Dong Yoon Choi, Dae Ha Kim, And Byung Cheol Song,” Recognizing Fine Facial Micro-Expressions Using Two-Dimensional Landmark Feature,” in 2018 25th IEEE International Conference On Image Processing (ICIP).

[122] Kai Xin Beh and Kam Meng Goh, “Micro-Expression Spotting Using Facial Landmarks,” in 2019 IEEE 15th International Colloquium on Signal Processing & its Applications (CSPA 2019), 8 -9 March 2019, Penang, Malaysia.

[123] Xiaohua Huang , Su-Jing Wang, Xin Liu, Guoying Zhao, Xiaoyi Feng , and Matti Pietik€ainen ,” Discriminative Spatiotemporal Local Binary Pattern with Revisited Integral Projection for Spontaneous Facial Micro-Expression Recognition,” In Ieee Transactions On Affective Computing, Vol. 10, No. 1, January-March 2019.

[124] Yuan Zong, Wenming Zheng, Zhen Cui, Guoying Zhao, and Bin Hu, ”Toward Bridging Microexpressions From Different Domains,” in IEEE Transactions on Cybernetics Volume: 50, Issue: 12, Dec. 2020.

LIST OF PUBLICATIONS

[1] The study performed in this work is submitted as "Micro-Expression Affect Recognition (MAR) : A Review and Recent Advances" for possible publication in “IEEE Transactions on Affective Computing”. The process for publication is under review.