Design and Analysis of an On-Chip Processor for Autism Spectrum Disorder Children Assistance Using Their Emotions

A Dissertation
Presented by

ABDUL REHMAN ASLAM
2017-06-0056

In Partial Fulfilment
of the Requirements for the Degree of
Doctor of Philosophy in
Electrical Engineering

Supervisor: **DR. MUHAMMAD AWAIS BIN ALTAF** (**LUMS**)

LUMS
A Not-for-Profit University

Syed Babar Ali School of Science and Engineering
Lahore University of Management Sciences
January 2023

Abstract

Autism Spectrum Disorder (ASD) is a "spectrum" neurological disorder causing different physical and cognitive disabilities. A major dilemma faced by ASD patients is the deregulated emotions causing certain, unpredicted, and instantaneous bursts of negative emotions. These negative emotion outbursts (NEOB) cause severe self-injuries and are a major hurdle for the treatment and rehabilitation of ASD patients. The unavailability of biomarkers for early prediction along with the life-long nature of the disorder requires life-long medical care and assistance for ASD patients. A physical or cognitive assistance system that can ease the severe difficulties faced by ASD children due to these certain NEOBs is direly required.

This Ph.D. thesis has targeted the early prediction of NEOB for ASD children. Early prediction can help the parents and caregivers control and regularize their emotions. I have proposed and developed a wearable system-on-chip (SoC) based digital back-end (DBE) processor for negative emotion and NEOB prediction using Electroencephalogram (EEG) signals. The feasibility of the wearable SoC processor is highly dependent on the size (area), battery life (energy), number of EEG electrodes, scalp location of electrodes for the patient's comfort, classification results, and the validation of the processor. A miniaturized, low-power SoC processor with a limited number of electrodes can be embedded in a headband as a patch sensor for continuous (24/7) prediction of NEOB/negative emotions. I have proposed and developed two (1st generation and 2nd generation) SoC-based DBE processors for the NEOB prediction. The identification of the most suitable channels and features for emotion prediction is an important challenge for researchers. I have also performed an extensive large-scale feature extraction using multiple benchmark emotions prediction data sets to identify the most suitable channels and features.

The 1st generation processor (DBE^1) and 2nd generation processor (DBE^2) provide the negative emotions prediction with 73.4% and 85.4% classification accuracy respectively. The DBE processors were designed and developed using a 180nm CMOS process. The DBE^1 processor was implemented using only eight EEG channels, whereas the DBE^2 was implemented using only two (minimum) channels. The DBE^1 and DBE^2 utilize an area of 5.4 mm^2 and 16 mm^2 respectively. The energy utilization of DBE^1 and DBE^2 is 16μJ and 10.13μJ per prediction respectively. The temporal and frontal location EEG electrodes are used in both DBE processors. ASD children are highly sensitive. Therefore, the location of the EEG electrodes is very important. The temporal and frontal locations cause minimum discomfort for the patients. Both DBE processors were validated using SEED and DEAP emotion prediction data sets.

Declaration

I, the undersigned, declare that this dissertation has been compiled by myself and its contents have not been presented for a higher degree at any other institute. The work in this dissertation is a result of original research carried out by myself under the supervision of Dr. Muhammad Awais Bin Altaf (Lahore University of Management Sciences, Pakistan), in partial collaboration with Prof. Dr. Hadi Heidari (University of Glasgow, Scotland, U.K). Most of the work in this dissertation has either been published as journal papers, conference proceedings, PhD forum, and book chapter, or is currently under review. A list of research publications and awards during this PhD is given on next page.

A significant portion of this research is my own work.

A Rehman

Abdul Rehman Aslam

Department of Electrical Engineering,

Syed Babar Ali School of Sciences and Engineering,

Lahore University of Management Sciences,

Lahore, Punjab, 47592, Pakistan.

Research Highlights

The Ph.D. topic is published in multiple top IEEE conferences, journals, and a book chapter. My Ph.D. work was also selected for multiple international and national research awards and grants. Following is the list of my awards and achievements:-

Awards

1 IEEE Circuits and Systems Pre-Doctoral Award, 2021.

2 2022 IEEE Design, Automation Test in Europe Conference Exhibition (DATE) PhD Forum.

3 Syed Babar Ali Research Fellowship Award, LUMS, Pakistan, 2020-22.

4 Commonwealth Split-Side Ph.D. Fellowship, University of Glasgow, UK, 2021-22.

5 IEEE Circuits and Systems Student Travel Grant, 2022.

Journals/Transactions

1 **<u>Abdul Rehman Aslam</u>**, Nauman Hafeez, Hadi Heidari, and Muhammad Awais Bin Altaf, "Channels and Features Identification: A Review and a Machine-Learning Based Model With Large Scale Feature Extraction for Emotions and ASD Classification." ***Frontiers in Neuroscience (FRNS)***, vol. 16 844851. July. 2022.

2 **<u>Abdul Rehman Aslam</u>** and Muhammad Awais Bin Altaf, "A 10.13 μJ/Classification 2-Channel Deep Neural Network Based SoC for Negative Emotion Outburst Detection of Autistic Children," in ***IEEE Transactions on Biomedical Circuits and Systems (TBioCAS)***, vol. 15, no. 5, pp. 1039-1052, Oct. 2021.

3 **<u>Abdul Rehman Aslam</u>** and Muhammad Awais Bin Altaf, "An On-Chip Processor for Chronic Neurological Disorders Assistance Using Negative Affectivity Classification," in ***IEEE Transactions on Biomedical Circuits and Systems (TBio-CAS)***, vol. 14, no. 4, pp. 838-851, Aug. 2020.

4 Muhammad Sheeraz, **Abdul Rehman Aslam**, Hadi Heidari, and Muhammad Awais Bin Altaf, " A Closed Loop Ear Wearable EEG Measurement Device with Real-time Passive Electrode Skin Impedance Measurement for Autism Prediction," in *IEEE Transactions on Circuits and Systems 1 (TCAS-1)*, **2022. [Under Review]**

Book Chapters

1 **Abdul Rehman Aslam** and Muhammad Awais Bin Altaf, "Machine learning–based patient-specific processor for the early intervention in autistic children through emotion detection," *Neural Engineering Techniques for Autism Spectrum Disorder*, Chapter 14, pp. 287-313, July. 2021.

Conferences

1 **Abdul Rehman Aslam** and Muhammad Awais Bin Altaf, "An 8.62 μW Processor for Autism Spectrum Disorder Classification using Shallow Neural Network," *2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, 2021, pp. 1-4.

2 **Abdul Rehman Aslam**, Talha Iqbal, Mahnoor Aftab, Wala Saadeh and Muhammad Awais Bin Altaf, "A 10.13 μJ /classification 2-channel Deep Neural Network-based SoC for Emotion Detection of Autistic Children," *2020 IEEE Custom Integrated Circuits Conference (CICC)*, 2020, pp. 1-4.

3 Mahnoor Aftab, Syed Adeel Ali Shah, **Abdul Rehman Aslam**, Waala Saadeh and Muhammad Awais Bin Altaf, "Design of Energy-Efficient Electrocorticography Recording System for Intractable Epilepsy in Implantable Environments," *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2020, pp. 1-5.

4 **Abdul Rehman Aslam** and Muhammad Awais Bin Altaf, "An 8 Channel Patient Specific Neuromorphic Processor for the Early Screening of Autistic Children through Emotion Detection," *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2019, pp. 1-5.

Acknowledgments

بِسْمِ اللهِ الرَّحْمٰنِ الرَّحِيمِ

"All Praise be to God, Lord of the Universe(s) "

"O Allah, send your grace, honor, and mercy upon Muhammad (peace be upon him) and upon the family of Muhammad (peace be upon him) "

Adoration and high esteem for my dear **parents** who have been a supporting shadow all these years, whose love, guidance, and prayers have made this day possible.

Completing Ph.D. studies can be compared to getting some food, after a long hunger. but for me, it was a fully guided journey with continuous refreshments. For this, I have the following people to thank.

First, I would like to express my deepest gratitude to none other than my supervisor **Prof. Muhammad Awais Bin Altaf**. I feel very lucky for being a Ph.D. student of Prof. Altaf. His continuous encouragement, support, rigor, and mentorship enabled me to complete this research work. He is an extraordinary mentor who guides his students from scratch and provides exceptional support in all aspects. I am thankful to my ex-student and colleague, Syed Muhammad Abubakar who introduced me to Prof. Altaf.

Profound regards to **Prof. Hadi Heidari** for his guidance and support during and after my split-side Ph.D. award for the University of Glasgow, UK. Although the research was carried out online due to Covid-19 restrictions, but he provided his maximum support and guidance. I am also thankful to my PhD committee members **Dr. Muhammad Adeel Ahmed Pasha** and **Dr. Nadeem Ahmad Khan** for their guidance.

I would also like to thank my friends and colleagues in the research group, that I have been a part of during my doctoral studies at LUMS, for providing a friendly and supportive atmosphere.

Deepest appreciation to **LUMS** and **Syed Babar Ali** for providing me with this wonderful opportunity. Also, I would like to thank the university for supporting my research

Contents

List of Figures

xvi

List of Tables

Abbreviations and Acronyms

3DCNN	Three dimensional convolutional neural network
ACCM-ABS	Accumulator and Absolute unit
AFE	Analog Front End
AIHPD	Absolute Interhemispheric Power Difference
A-LSTM	Attention Long Short-Term Memory
ANN	Artificial Neural Network
ASD	Autism Spectrum Disorder
ASKI	Approximated Skewness Indicator
BM-DS	Benchmark Data Sets
CFU	Classification Unit
CLF	Classification Algorithm
CSP	Common Spatial Pattern
DASM	Differential Asymmetry
DBE	Digital Back End
DBE^1	1st Generation Digital Back End Processor
DBE^2	2nd Generation Digital Back End Processor
DBS	Deep Brain Stimulation
DCAU	Differential Caudality
DNN	Deep Neural Network
ECG	Electrocardiogram

EDL	Emotion Decision Logic
EEG	Electroencephalogram
FEE	Feature Extraction Engine
FN	False Negative
FNU	Feature Normalization Unit
FP	False Positive
FLP	Floating Point
GELM	Graph regularized Extreme Learning Machine
HHT	Hilbert-Huang transform
HJ	Hjorth
HL	Hidden Layer
HOC	Higher Order Crossings
HVHA	High Valence High Arousal
HVLA	High Valence Low Arousal
IL	Input Layer
iEEG	intracranial EEG
IHPR	Interhemispheric Power Ratio
INC	Increment
KNN	K Nearest Neighbour
LIHPR	Logarithmic Interhemispheric Power Ratio
LR	Logistic Regression
LSFE	Large Scale Feature Extraction
LVHA	Low Valence High Arousal
LVLA	Low Valence Low Arousal
ML	Machine Learning

MMG	Mechanomyogram
MUX	Multiplexer
NEOB	Negative Emotion Outburst
NFV	Normalized Feature Vector
OL	Output Layer
PGA	Programmable Gain Amplifier
PSD	Power Spectral Density
PPU	Pre-Processing Unit
RASM	Rational Asymmetry
RBF	Radial Basis Function
RLU	RELU Activation
SGM	Sigmoid Activation Function
SKEW	Skewness
SoC	System On Chip
STFT	Short Term Fourier Transform
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
ULP	Ultra Low Power
ZCD	Zero Crossings Detector

Chapter 1

Introduction

Autism spectrum disorder (ASD) is a wide spectrum of various neurological ailments that include an amalgam of different disorders along with one or multiple disabilities. The disabilities include cognitive weaknesses, behavioral, and communication impairments, etc. A significant ratio of ASD patients also suffer from other neurological disorders, e.g. epilepsy, developmental delay, tuberous sclerosis, and fragile X syndrome [3]-[4]. Therefore, ASD is termed as a "**spectrum**" disorder and considered as the most complex neurological disorder because of wide range of symptoms, complications and severity [5].

The recent statistics of ASD are quite alarming. Center for disease control and prevention network's report for ASD prevalence in 2021 has revealed that 1 in every 44 children is suffering from ASD based on 2018 data, compared to 1 out of 54 children in 2016 [6]. There is an alarming 241% increase in the number of ASD cases from 2000 to 2018 [6]. Figure 1-1 shows the number of cases per million from 2000 to 2020 which highlights the sharp increase. The prevalence of ASD in underdeveloped countries is far more concerning. Pakistan, for instance, lacks reliable ASD statistics and is mostly unaware of the disorder and its consequences [7]. It is reported that 0.35 million children in Pakistan are affected by this disorder [8]. The real statistics may be much worse because of unawareness about the disorder, and the unavailability of basic health facilities to a wide ratio of the population.

The most challenging factor about ASD is the physical impairments and intellectual disabilities associated with the disorder. These disabilities and the difficulties have significant ramifications on the quality of life of ASD patients and their caregivers, resulting in stress,
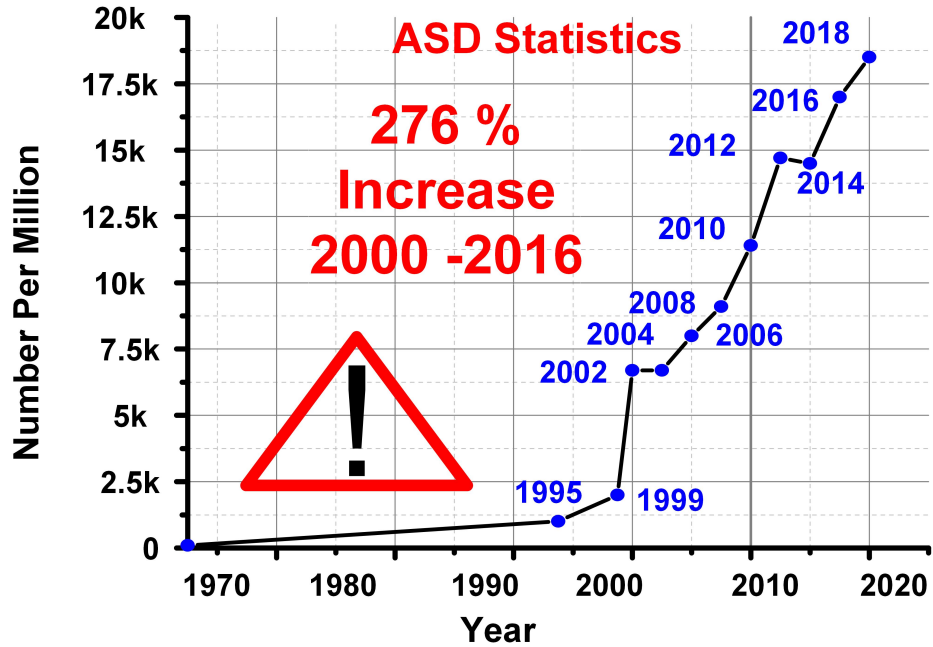
Figure 1-1: Center for Disease Control & Prevention Report 2021, Autism Spectrum Disorder Statistics in the United States of America.
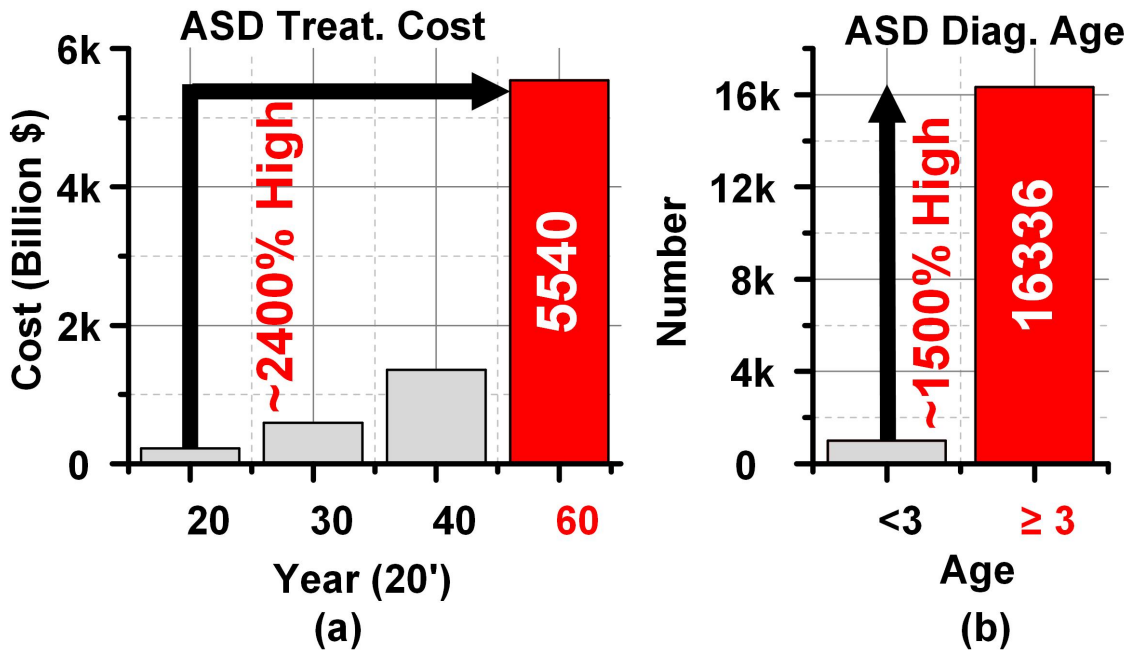
Figure 1-2: (a) Autism Spectrum Disorder Treatment Cost, United States of America (b) Autism Spectrum Disorder Diagnosis Age, Scotland.

emotional issues, and financial burden. The physical impairments involve certain motor disabilities causing involuntary hand and leg movements, eye movements, etc. [9]. The intellectual disabilities include a low intelligence quotient level, which creates difficulties in employment, behaviors, and social relationships. The large patient count with related physical and mental disabilities is creating a huge financial impact on the health budget. The cost of treatment for ASD patients is projected to reach 5.54k billion USD in the United States of America in 2060 [10]. Figure 1-2 (a) shows the estimated ASD treatment budget in the United States of America from 2020 to 2060. It can be seen that the projected annual ASD treatment budget in 2060 is $\approx 2400\%$ higher than in 2020.

Diagnosing autism is also a challenging process due to the unavailability of medical and pathological diagnosis tests [11]. That's why late diagnosis and hence late treatment is a major problem for ASD patients. The recent age-related statistics for ASD diagnosis also reveal that a majority of ASD patients are diagnosed at later ages. Figure 1-2 (b) shows the age statistics of ASD patients in Scotland [12]. It was observed that the number of ASD patients diagnosed at later ages is quite higher as compared to the earlier ages. The number of diagnosed cases in age ≥ 3 years is 1500% times higher than age < 3 years. The current ASD diagnosis methods require extensive developmental and behavioral screening by neurologists. Autism Diagnosis Observation Schedule- Second Edition (ADOS-2) is a standard ASD diagnosis and assessment method [13]. ADOS-2 involves extensive behavioral evaluations to evaluate ASD markers. The neurologists perform a list of behavioral evaluations, e.g. communication index, reciprocal social interaction, imagination and creativity, and stereotyped behavior score for each suspected patient, and compare the results with a cut-off table to mark the investigated patient as an ASD patient.

1.1   Emotional Disorders in Autsim

The most demanding factor related to ASD patients is their lifelong cognitive disabilities associated with the disorder and coping with the irregular emotions caused due to these disabilities. It has been observed that ASD patients have significant emotional impairments leading to depression and anxiety disorders [14]. The emotional impairments produce volatile emotions and emotional deregulation [15]. Emotion deregulation is described as the failure to regulate emotions appropriately and effectively. The emotional deregulation cre-

ates maladaptive emotional responses leading to anger control problems, temper tantrums, and aggression [16]. These maladaptive emotional responses cause physical damage to self and others. It has been observed that ASD patients are highly affected by these emotional disorders and swift negative emotions [17]. The primary reasons for this negative esteem of emotions and emotion deregulation are the mistreatment and oppression in schools, non-flexible society behaviors, stigma associated with their cognitive deficiencies, physical and sensory deficiencies, and the inability of parents and teachers to understand their state of mind.
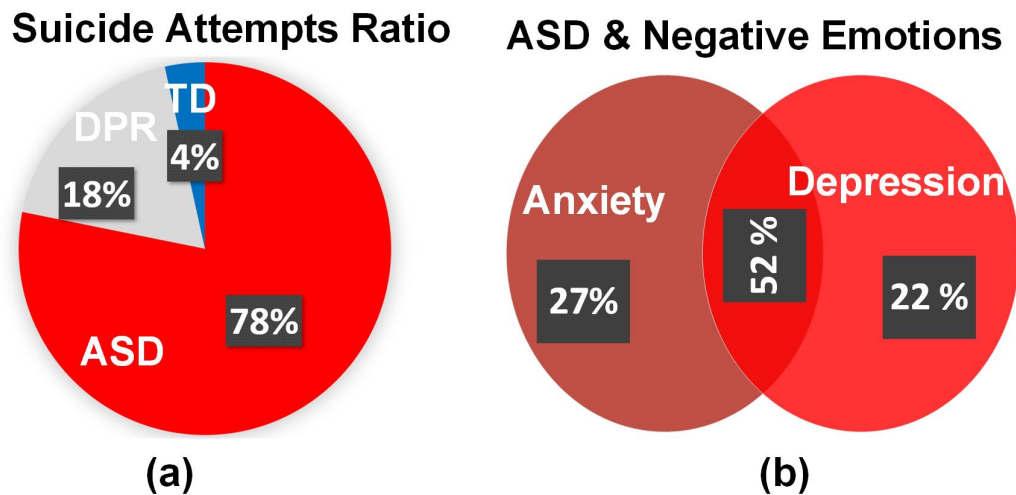
Figure 1-3: (a) Suicide Attempts in Autism Spectrum Disorder (b) Anxiety & Depression in Autism Spectrum Disorder.

These negative, uncontrolled, and deregulated emotions, are a primary reason for the large ratio of suicide attempts reported in ASD (78.16%) and depressive (DPR) children (18.38%) in comparison to the typically developing children (3.46%) as depicted in Fig 1-3 (a) [18]. It is also reported that a large number of ASD patients are either suffering from anxiety, depression, or both anxiety and depression, as shown in Figure 1-3 (b) [17]. The early recognition of the anger triggers is important to avoid these maladaptive and deregulated emotions. Therefore, it is vital to predict the emotions of ASD patients to regulate their emotions and avoid self-injuries [15].

Emotion prediction is a serious deficiency and pain problem in ASD children. The ASD patients cannot identify their emotions timely and correctly [19]. The inability of timely prediction of emotions leads to unregulated emotions. Different solutions are used

to identify and then regulate the emotions for ASD patients. The solutions include emotion identification using emotion wheels and cards, or surgical solutions.

The ASD children are trained to express their emotions by pointing out to the corresponding place on the emotions wheel or picking the equivalent emotion card in the emotional wheels and card mechanism. However, this is not reliable because of the inability of early prediction, and the lack of self-control among ASD children. The surgical solutions include deep brain stimulation (DBS) for ASD patients with serious emotional issues [20]. Electroencephalogram (EEG) electrodes are implanted invasively in the patient's brain. Electrical pulses are transmitted inside the brain using a neuro-stimulator. The location of electrodes and the stimulation effects different brain functions. Severe emotions are managed by stimulating specific parts of the human brain in clinical settings [21].

The limited DBS studies have shown positive results for emotional disorders. But they have severe limitations. There are chances of misplacement of leads for DBS which can cause bleeding inside the brain, strokes, infection, headache, and seizures [21]. The primary reason for these side effects is the invasive nature of the procedure. Small holes are created in the skull for electrode implantation. The DBS process also has extensive clinical settings, trained medical staff and patient admission to some hospitals. The 24/7 access of these facilities cannot be provided to each ASD patient, especially in third-world countries, poor income groups, and undeveloped areas. Therefore, a noninvasive, wearable, and closed-loop system to predict and control the emotions for ASD patients is direly required.

1.2   Closed Loop Emotion Prediction System

The early identification of emotions is a key to the identification of self-harming negative emotions to avoid injuries and regulate the emotions. The ASD children have reported irrepressible mental feelings before the self-harming negative emotions [22]. This thesis targets to use these irrepressible mental feelings for the earlier prediction and control of these negative emotions. The EEG signals of the ASD patients are used for this negative emotions prediction system. Figure 1-4 depicts the broad idea of the proposed system to support an ASD child during specific learning activities. The top marked as red shows the state of an ASD child without any emotional prediction system. The patient's mind is surrounded by negative emotions as portrayed in the figure. Since the ASD child is

affected by limited social syndrome and unable to express his distress and negative emotions, therefore fails to get any benefit from the learning activity. The bottom part (blue dotted) shows the proposed effective intervention system which will extract the EEG data, process the information, and assist the caregiver in molding the learning activity tuned to the specific patient's needs after necessary emotional prediction and control. A closed-loop and child-specific audio/video stimulation can be provided to avoid negative emotions and relax ASD children. The presented figure exemplifies how the proposed device with real-time human emotion classification can play a vital role in coping with social interaction issues for ASD children.
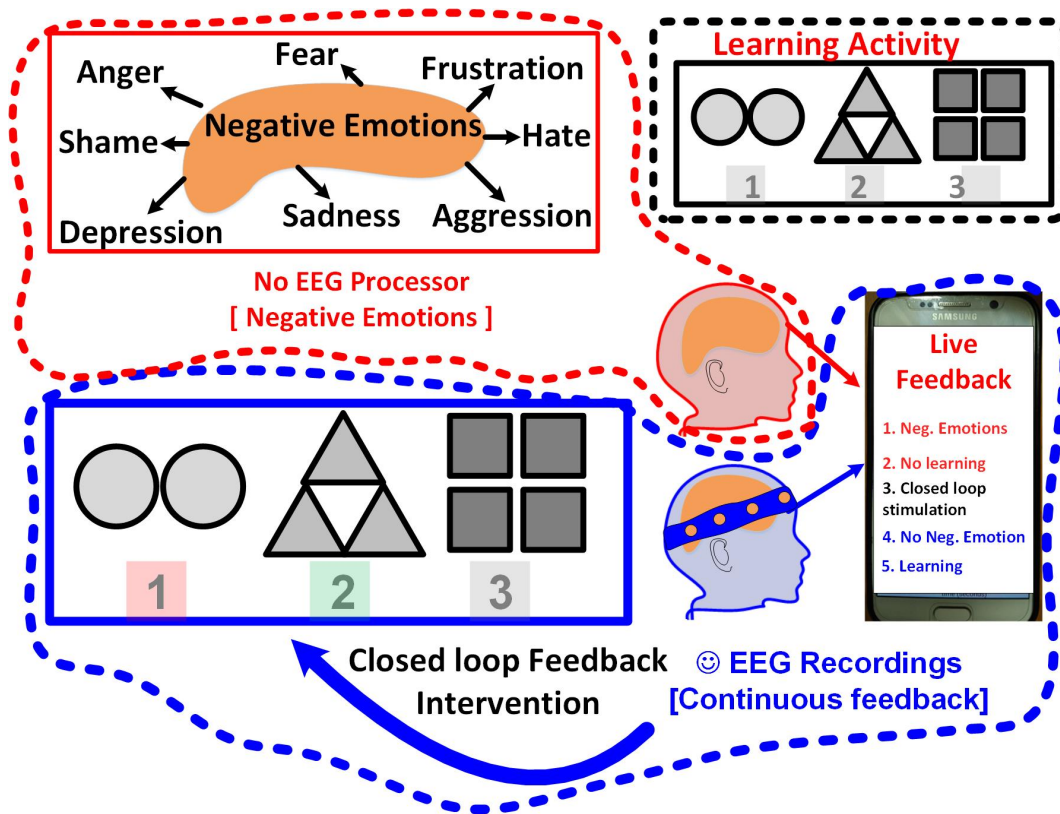
Figure 1-4: Proposed Autism Spectrum Disorder Negative Emotion Prediction & Closed-Loop Stimulation Solution.

A miniaturized, fully integrated, and wearable system-on-chip (SoC) based emotions prediction processor is used for this closed-loop emotions control system. The complete emotions prediction SoC involves the design and development of an analog front end (AFE) and a digital back end (DBE) processor. The AFE is used for real-time data acquisition and the DBE is used for emotions prediction using the data acquired through AFE. This Ph.D.

thesis targets a miniaturized, ultra-low power (ULP), and fully integrated DBE processor for negative emotions prediction and control.

1.3   Negative Emotions Prediction DBE Processor

Human emotion recognition can be performed through 1) physical signals i.e. facial expression, speech, and posture, or 2) physiological signals, i.e. EEG, electrocardiogram (ECG), and temperature. Physical signals are easy to collect but lack reliability. On the other hand, physiological signal tracking does not have reliability issues but acquiring the data can be a bottleneck. The EEG signals are highly suitable for the design and development of a wearable on-chip emotions classification processor [23]. Therefore, I have used the EEG signals for the design and development of the wearable SoC-based emotions prediction DBE processor.

Figure 1-5 shows the overall flow of the EEG-based emotions prediction and control system for ASD patients. The EEG signal acquisition is performed using AFE and ADC. The real-time EEG data is forwarded for feature extraction and classification for negative emotions prediction. A closed loop feedback is provided to control the emotions using a suitable intervention.
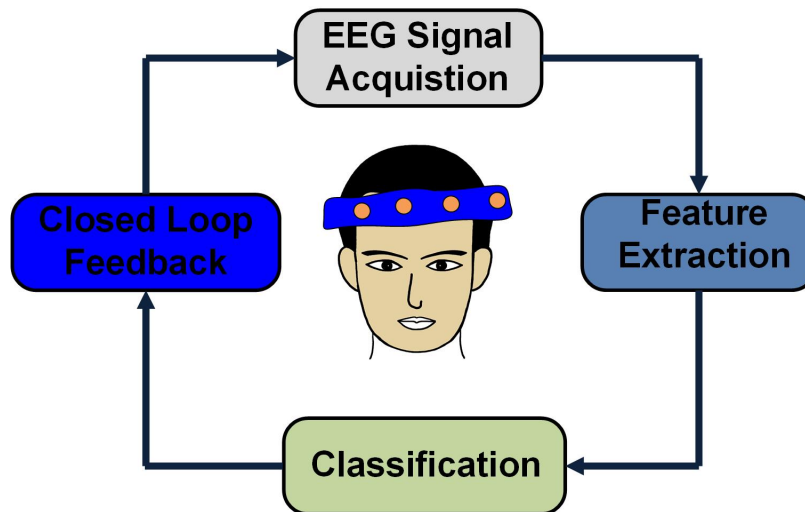
Figure 1-5: Scalp EEG Based Emotions Prediction & Control for Autism Spectrum Disorder.

The DBE emotions prediction processor developed in this Ph.D. project is targeted for a

wearable SoC that can be embedded inside a headband as depicted in Figure 1-5. The DBE operates with a limited number of channels for a miniaturized SoC with low classification energy, and early prediction (low latency), and is operable with a coin cell battery of size, area, voltage, and current of 20 mm, 3 V, and 250 mAH, respectively.

1.4 EEG Emotions Prediction: Previous Solutions

To the best of my knowledge, no fully integrated SoC has been developed to predict the negative emotions before their physical effect using electrical onset and provide closed loop feedback to suppress them. Most of the previous emotions prediction systems are either 1) software based 2) FPGA-based prototype or 3) does not provide fully integrated SoC solution [24],[25], [26]. The software-based models are not suitable for a 24/7 wearable device. The FPGAs are used as a prototype for the SoC testing and cannot provide a wearable implementation [25]. Some partially integrated SoCs have been developed. But they only provide on-chip classification and lack feature extraction.

1.5 Proposed Solution: Challenges and Contributions

The long-term monitoring of the EEG patterns is required for the analysis of the neurologists for ASD patients to track irregular negative emotions. However, the power overhead due to wireless communication can abruptly drain the battery. It would also continuously dissipate heat causing discomfort for ASD patients. However, fully on-chip emotions prediction using suitable features and classification algorithms would reduce the power overhead. The EEG patterns causing negative emotions can be recorded for later analysis by neurologists. The full integration of feature extraction and classification algorithms on-chip for wearable devices reduces the power consumption of the system by 80X [27]. However, the design and development of an SoC-based DBE processor for negative emotions prediction has the following main challenges:-

- **Number of Channels:** Each EEG electrode provides electrical activities for a specific part of the human brain. That's why many state-of-the-art works with high classification results utilize a maximum number of channels. The primary focus of these works are the classification results. However, fully on-chip wearable SoC systems

8

have very strict area, power, classification energy, and patient comfort constraints. It is highly challenging to achieve high classification results with limited channels on specific locations (frontal, temporal) causing minimum discomfort.

- **Validation:** Many high accuracy emotions prediction systems are validated on a limited number of subjects and private data sets. However, the design and development of an SoC-based negative emotions prediction system involve a significant cost. The systems validated on a limited number of subjects and local data sets may fail to generalize for benchmark data sets. Therefore, wearable SoC systems should be validated on multiple benchmark data sets with a large number of subjects before physical implementation.

- **Area, Power and Classification Energy:** Many state-of-the-art works utilize highly complex features and classification algorithms. These features and classification algorithms cannot be implemented for a low-power and energy SoC. It is highly challenging to find low-complexity features and classification algorithms with the maximum classification results. The high-complexity features and classification algorithms require extensive optimizations and tuning for a low area, power, and classification energy implementation.

- **Fully Integrated DBE Processor:** There are some SoC-based solutions for emotion prediction with offline feature extraction. The features are extracted offline and then forwarded to the emotions prediction chip. The offline feature extraction does not fulfill the requirements for a wearable SoC solution. Therefore, a fully integrated SoC with feature extraction and classification units is required.

- **Large Scale Feature Extraction Analysis:** Selection of the minimum number of EEG channels is a very challenging problem, and the selection is based on experimentation of a large feature set and literature review. I have presented the first large-scale feature extraction analysis for emotion prediction. The large-scale feature extraction analysis points out the most suitable channels and features for emotion prediction.

This dissertation presents the design and analysis of a DBE processor for a miniaturized, energy-efficient, and wearable SoC for monitoring, prediction, and suppression of negative emotions with a minimum number of electrodes. Figure 1-6 highlights the shortcomings in

Previous Solutions

- Invasive Solutions
- Software-Based Solutions
- Low Classification Results ☹
- Large No of Channels
- Limited Pvt Data Set Validation
- Large Area
- High Classification Energy
- Lack Full Integrated Solution

1ˢᵗ Generation DBE

- SoC Based Solution
- Linear SVM Classifier
- High CLS Results
- Low # of Channels
- BM-DS Validation
- Miniaturized Area
- Low CLS Energy
- Fully Integrated

Implemented Results	
Process	0.18um 1P6M CMOS
Area	FE: 3.6mm² CLS: 1.4mm² SVC: 0.4mm²
Supply	1.0 V
# of Channel	Eight (8)
CLS	Linear SVM
Power/ Energy	2.04mW/ 16uJ/CLS
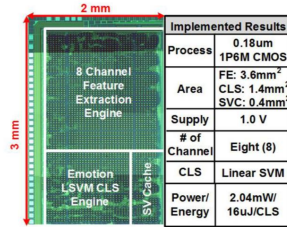
This Work

- Non Invasive Wearable Solution
- SoC Based
- High Classification Results ☺
- Low No of Channels
- Multiple Data Sets Validation
- Low Area
- Low Classification Energy
- Fully Integrated Solution

2ⁿᵈ Generation DBE

- SoC Based Solution
- DNN Classifier
- Excellent CLS Results
- Min # of Channels
- BM-DS Validation
- Miniaturized Area
- Min CLS Energy
- Fully Integrated

Implemented Results	
Process	TSMC 0.18um 1P6M CMOS
Area	4.0 x 4.0 mm
Supply Voltage	1.0V (AFE) 1.0V (DBE)
NEF/channel	2.71
Classifier	DNN
Energy Efficiency	10.13uJ/Class.
Accuracy	85.2%
No. of Classes	4
PS Weight Memory	16KB
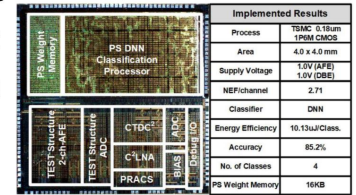
Figure 1-6: Shortcomings in Previous Solutions and Contributions of My Work.

previous solutions targeting negative prediction, suppression, and control of negative emotions. To the best of my knowledge, no previous work has targeted a non-invasive, wearable, and fully integrated SoC-based solution with high classification results, a minimum number of EEG electrodes with the low area, and classification energy solution, validated on multiple benchmark data sets for negative emotions prediction and control. To achieve this target, I have designed and developed the 1st and 2nd generation DBE processors, named DBE^1 and DBE^2, respectively for negative emotions prediction as highlighted in Figure 1-6. The DBE^1 uses a linear support vector machine (SVM) classifier to provide negative emotions prediction with a small number (8) of EEG channels and high classification accuracy (73.14%). The DBE^2 uses a deep neural network (DNN) classifier with a minimum number (2) of EEG channels to provide negative emotions prediction with a maximum

classification accuracy (85.4%). Both the DBE processors were fabricated using TSMC 180 nm CMOS process and validated using multiple benchmark data sets (BM-DS) for emotion prediction.

The design and development of DBE^1 and DBE^2 involved different challenges. Some of these challenges and the proposed solutions utilized in this work are summarized below:-

- The Interhemispheric Power Ratio (IHPR) feature for DBE^1 requires complex floating point divider implementation, which was replaced and approximated with a Logarithmic interhemispheric Power Ratio (LIHPR) with an overall area reduction by 4.7X.

- The skewness feature for DBE^2 has a highly complex hardware implementation. The feature was replaced and approximated by an approximated skewness indicator (ASKI) with 86X reduced area.

- The conventional implementation for the DNN utilized in DBE^2 required a huge area and does not suit an ULP implementation. The DNN was implemented in a semi-pipe-lined manner with 77% reduced energy.

- The positive and negative symmetry of the sigmoid function was utilized to implement the sigmoid unit with 50% lesser memory in DBE^2.

1.6   Thesis Organization

This organization of this Ph.D. thesis is presented in Figure 1-7. The chapter-wise summary of this thesis is presented below:-

- **_Chapter 1:_** The challenges faced by ASD patients are explained and the reasons for selecting this topic are highlighted in Chapter 1. The statistics of ASD including the number of patients, diagnosis age, challenges in ASD diagnosis, and the emotional disorders faced by ASD patients are also discussed in this chapter. The reasons for the requirement and the need of a closed-loop assistance system using emotions for ASD patients are explained in this chapter. How an SoC-based solution using EEG signals can be used to control emotional disorders, and assist ASD patients is discussed in chapter. The challenges and contributions of this PhD work to design a DBE processor for negative emotions prediction are also highlighted in this chapter.
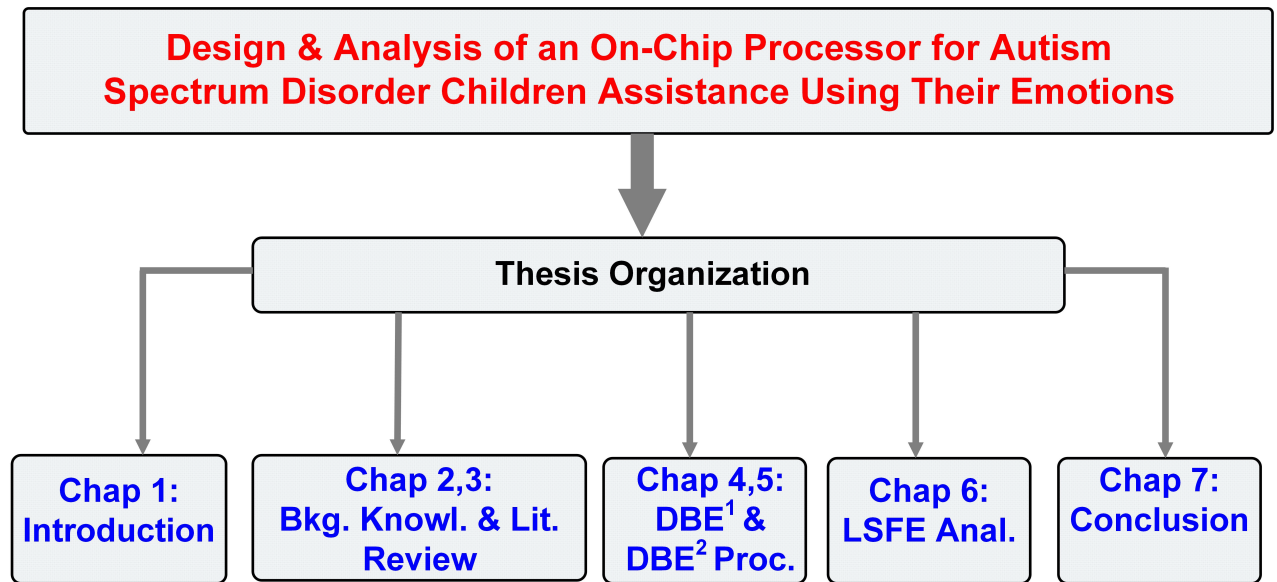
Figure 1-7: Thesis Organization.

- **_Chapter 2:_** The necessary background knowledge for the readers of this thesis is provided in Chapter 2. A brief overview of human emotions, measuring human emotions, EEG signals and their electrode placement, EEG-based emotions prediction data sets, and labels in emotions prediction are provided. The utilization of machine learning for emotion prediction with a cloud computing-based approach or a fully on-sensor SoC-based system is discussed. The design challenges associated with the design and development of the DBE processor for emotion prediction are discussed in this chapter.

- **_Chapter 3:_** provides an organized literature review of the previous state-of-the art works. The comparison metrics of this work including hardware implementation, fully integrated solution, benchmark data set validation, the number of EEG channels, feature vector, classification algorithm, area, and power/energy of emotions prediction processors are defined.

- **_Chapter 4:_** The design and analysis of the DBE^1 emotions prediction processor is explained in Chapter 4. The DBE^1 was the 1st fully integrated SoC-based DBE processor for emotions prediction. The DBE^1 was designed and developed using a linear SVM classifier with power spectral density (PSD), LIHPR, and absolute inter-hemispheric power difference (AIHPD) features using eight EEG channels. The nega-

tive emotions prediction was performed using valence and arousal classification, using DEAP and SEED data sets. The proposed feature vector was implemented in an area and energy-efficient manner with 4.7X lesser area than conventional implementation. The DBE^1 provided the classification performance of 73.14% and 75.8% accuracies using DEAP and SEED data sets respectively. The DBE^1 was implemented using 180 nm TSMC CMOS process with an area of 5.4 mm^2. The classification energy consumed by DBE^1 was $16\mu J$ per prediction.

The selection of channels, features, and classification algorithm are explained in this chapter. The hardware architecture and optimization of the selected features for the minimum area and classification energy are also explained in this chapter. Finally, the classification performance of the DBE^1, and its comparison with the state-of-the-art works is provided in this chapter.

- **_Chapter 5_**: The design and analysis of the $DBE2$ for emotions prediction is explained in Chapter 5. The DBE^2 was designed and developed using a fully connected DNN classifier with zero crossings detector (ZCD), and skewness (SKEW) feature for two EEG channels. The proposed hardware implementation of the selected features utilized 86X lesser area as compared to the conventional implementation. A semi-pipe-lined DNN classification unit was implemented to reduce the classification energy by 77%. The $DBE2$ was also validated using DEAP and SEED data sets and fabricated using 180 nm TSMC CMOS process. The $DBE2$ utilized an area of $16mm^2$. The energy consumption of DBE^2 was $10.13\mu J$ per classification.

  The selection of channels, features, and classification algorithms are explained in this chapter. The hardware architecture and optimization of the selected features for the minimum area and classification energy are also explained in this chapter. Finally, the classification performance of the DBE^2 and its comparison with the state-of-the-art works is provided in this chapter.

- **_Chapter 6_**: The large-scale feature extraction (LSFE) methodology analysis for ASD and emotions prediction using maximum benchmark data sets is explained in Chapter 6. An extensive analysis using LSFE was performed to find the most suitable channels and features for ASD and emotions prediction. A comprehensive LSFE was carried out for the most suitable EEG channels and brain area identification in ASD and

emotions prediction. The identified brain area outperformed many state-of-the-art using a lower number of channels for emotions and ASD classification. The temporal, frontal, and prefrontal brain regions are included in the identified brain areas. A set of features highly suitable for emotions and ASD prediction, identified through LSFE is also provided in this chapter. The area and energy-efficient hardware implementation of these features can be used for hardware-based systems,

- *Chapter 7*: The overall conclusion of this Ph.D. work with the list of achievements, and the future work's directions are discussed in Chapter 7. The DBE processors designed and developed during this Ph.D. work are presented at multiple top international conferences including IEEE AICAS, IEEE ISCAS, and IEEE CICC. This This work was selected for multiple national and international awards including SBARA, IEEE CAS Pre-Doctoral award, IEEE CAS student travel grant, and Commonwealth Split Side Ph.D. fellowship. This work was also published and under review in multiple high-impact journals, transactions, and a book chapter including IEEE TBioCAS, FRNS, TCAS-1, and Elsevier's Neural Engineering Techniques for ASD. The future work directions include SoC implementations of binary neural networks, ULP AFE development, and a data set collection for ASD patients including physiological signals, EEG, and Mechanomyogram (MMG) signals.

Chapter 2

Background Knowledge

This chapter discusses the background knowledge required for the design and development of an EEG-based emotions classification processor. In this chapter, human emotions and their quantization models, EEG signals, placement of EEG electrodes, EEG classification process, and limitations for the design and development of wearable EEG-based emotions classification processors are discussed.

2.1 What are Human Emotions?

Human emotions are conscious or unconscious feelings of the human brain. Different theories defining human emotions have acknowledged that "***Human emotions are a complex phenomenon and cannot be fully described through behavioral experiences***" [28]. Scientists have identified that different human emotions excite different neurons [28]. They are a combination of different neurological, behavioral, and psychological processes. The complete description of human emotions should incorporate the conscious or unconscious feelings, and the changes occurring in the human nervous system. The most challenging problem related to human emotions is the measurement and identification of correct emotions.

2.2 Measuring Human Emotions

The quantization of human emotions is a notorious problem. The number and scale of present human emotions is an important question that does not have any defined answer. Currently, the two popular methods used for emotion quantization include a discrete set of emotions and dimensional scales. The discrete emotional scale maps a human emotion into a discrete set of basic emotions ranging between 6 to 14, e.g. anger, fear, sadness, etc. [29]. Table 2.1 lists some popular discrete sets of emotions along with different emotions. The emotions of each participant are mapped to the nearest emotion among the discrete set of emotions on that scale. The emotion of each participant is labeled as anger, fear, disgust, sadness, surprise, or joy using Ekman's six basic emotions [29]. Similarly, the emotions of each participant are mapped to the eight, six, and nine emotions using Plutchik, Parrot, and Tomkin scales, respectively [30], [31], [32].

Table 2.1: Different Discrete Emotional Scales.

Reference	Emotions
Ekman [29]	anger, fear, disgust, sadness, surprise, joy
Plutchik [30]	anger, fear, sadness, disgust, surprise, anticipation, trust, joy
Parrot [31]	anger, fear, love, joy, sadness, surprise
Tomkin [32]	anger, disgust, distress, contempt, interest, fear, shame, joy, surprise

A major problem with the discrete set of emotions is the identification of mixed emotions or the emotions that overlap between the available discrete set of emotions [33]. The dimensional scale of emotions measures emotions in multiple dimensions including valence, arousal, dominance, liking, etc. [33]. Russel's valence arousal scale is the most frequently used dimensional model for emotion measurement. The valence defines the scale of positivity or pleasantness of emotion, whereas the arousal refers to the scale of the intensity of that emotion [34].

Figure 2-1 depicts some discrete basic emotions mapped using the valence arousal scale. It can be observed that the angry and annoyed are negative emotions. However, annoyed emotion has more intensity than angry emotion. The calm and excited and calm and positive emotions with similar valence, but different arousing intensities. The valence and

16

arousal are scaled in a discrete range from a minimum to maximum value. The increase in valence and arousal values is represented from left to right and bottom to top, respectively as shown in Figure 2-1.
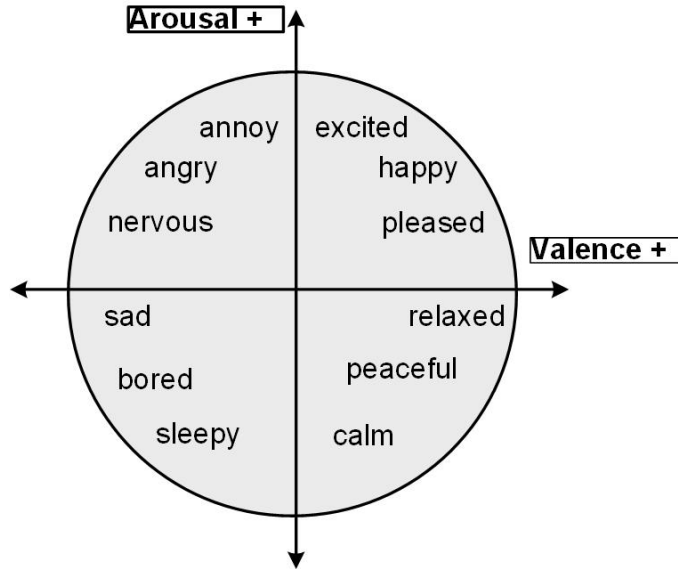
Figure 2-1: Valence Arousal Model & Basic Emotions.

2.3 Machine Learning

Machine learning (ML) is a branch of artificial intelligence, in which a machine (computer) targets to learn the structure of data and find a relationship between the output data (labels) and input data (features). ML is broadly categorized between supervised and Unsupervised learning. In supervised learning, the ML algorithm is provided with examples of the input and output values. The ML algorithm tries to establish a relationship between the inputs and outputs and then calculates the output values for unknown data using that relationship. Figure 2-3 depicts the process of a supervised ML algorithm using an example. In this example, different types of apples and guavas are provided to the ML model with their labels. The ML model learns to label an unknown example as an apple or guava. The ML prediction of an object is performed in two ways including classification or regression.
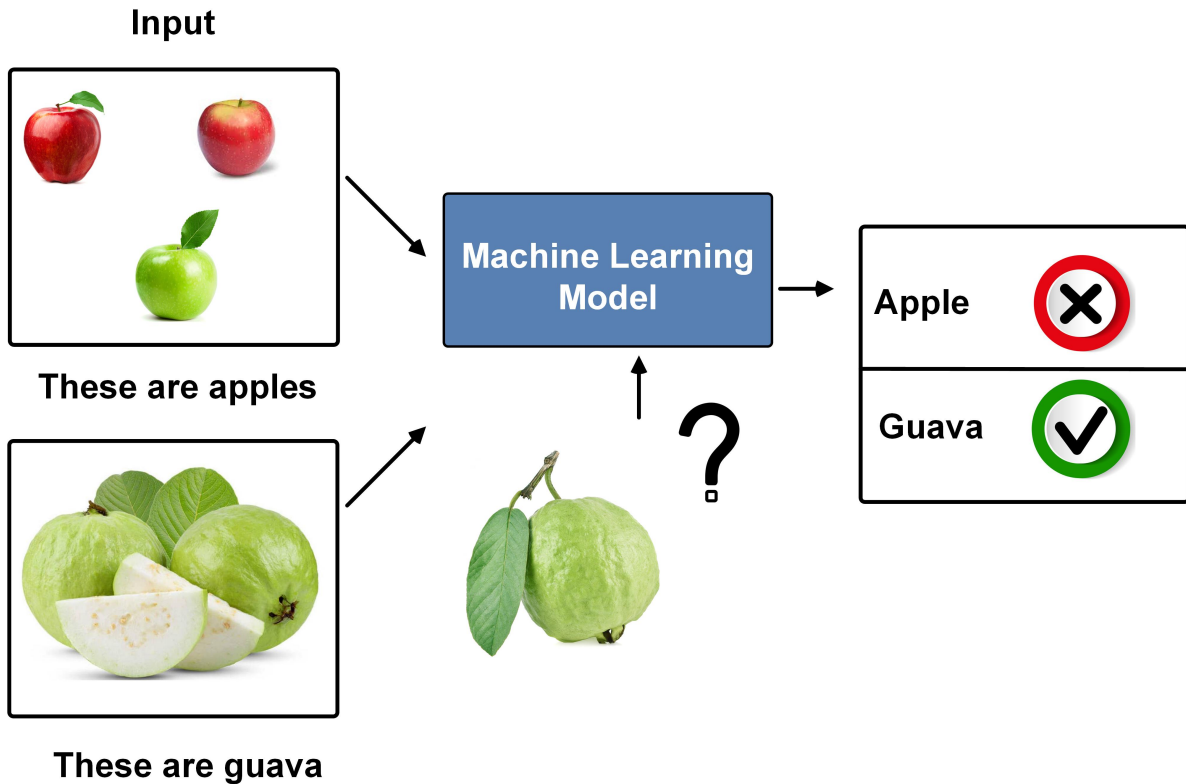
Figure 2-2: Supervised Machine Learning Example Demonstration.

2.3.1 Classification and Regression

An ML categorization of an object in a discrete set of labels is called **classification**, whereas the prediction of an object in some quantity or value is called **regression**. For e.g. the categorization of the fruits as apple or guava in the example discussed above would be classification. The prediction of the room's temperature in degrees Celsius or the value of a person's blood pressure would be regression. Since, in this Ph.D. project, my target was the prediction of negative emotions and ASD prediction, I have focused on supervised learning and ML classification.

2.3.2 Features Extraction and Classification Algorithm

The feature extraction process is the procedure to cut down the number of input values required for the output label prediction. For e.g in the above example of fruits predicted as an apple or guava, if the ML algorithm is provided with the input images of 500x500 resolution, then the ML algorithm would require to map 0.25 million pixels with the output

label. However, the features of texture smoothness, color, and shape (length, width, height) can be used for apple or guava prediction in much lower dimensions.

Classification algorithms in ML are the algorithms that utilize the input data (raw data or features) and map the output labels to the input data using some algorithm. The popular algorithms for ML classification are logistic regression (LR), naive Bayes (NB), K nearest neighbor (KNN), and SVM are the popularly used ML classification algorithms.

2.4 Electroencephalogram Signal

EEG signals are used to record the electrical activities inside the human brain. The human brain contains billions of neurons generating ionic currents for each neural activity. The cell membranes covering the brain neuron cells contain negatively or positively charged ions inside or outside. These electric charges flow through the calcium, potassium, magnesium, and sodium fluids in cell membranes. The neural activities generated by human brain cells cause small electrical currents to flow through these ionic channels. This electrical current is measured by placing electrodes inside the skull or on the scalp. The electrical current measured by placing electrodes inside the skull is called intracranial EEG (iEEG) [1] and the electrical current measured by placing the electrodes outside the scalp is called scalp EEG [2]. Since I was targeting a non-invasive wearable SoC solution. Therefore, I have focused on data sets utilizing scalp EEG.
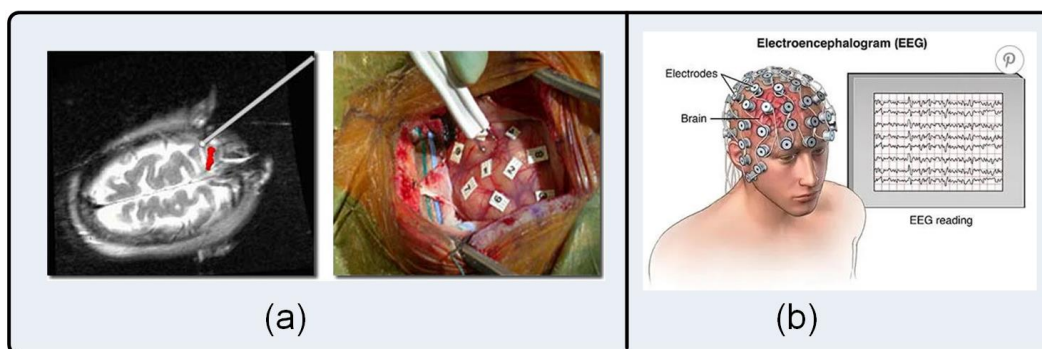
Figure 2-3: iEEG and Scap EEG (Pictures are taken from golbylab [1] and bright brain center [2].

Figure 2-4 shows a sample EEG signal of six seconds duration for a certain brain location. The electrical voltage measured in micro-volts (uV) is recorded on a specific sampling

frequency (128 Hz). The sampling frequency is the number of EEG samples recorded per second. A certain number of electrodes are placed on the human scalp to acquire EEG signals. The electrode placement position is proportional to the shape and size of the human skull.
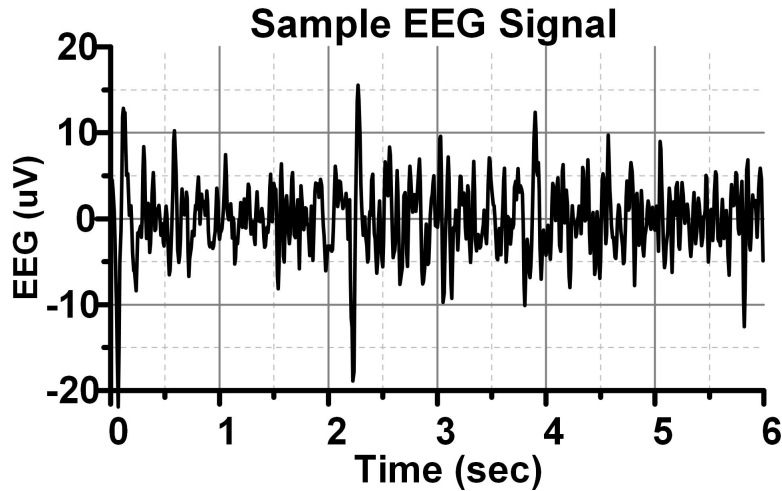
Figure 2-4: EEG Signal Sample.

The 10-20 EEG placement is the international standard method for the uniform placement of EEG electrodes. Figure 2-5 shows the placement of 32 EEG electrodes using a 10-20 layout. The EEG electrodes on the left side of the human head are highlighted in red color.

The 10% and 20% are used to represent the distance between electrodes [35]. The nasion refers to a point between the forehead and nose. The inion refers to a bump between the skull's bottom and neck [35]. The EEG electrodes record the brain signals of all regions including frontal, parietal, temporal, and occipital lobes represented by letters F, P, T, and O, respectively. The EEG electrodes covering right and left hemisphere locations are numbered with odd and even numbers, respectively. The EEG electrodes covering the midline of the hemisphere are numbered with Z. For example Fz, F3, and F4 represent the midline, left side, and right side frontal lobe electrodes, respectively.

20
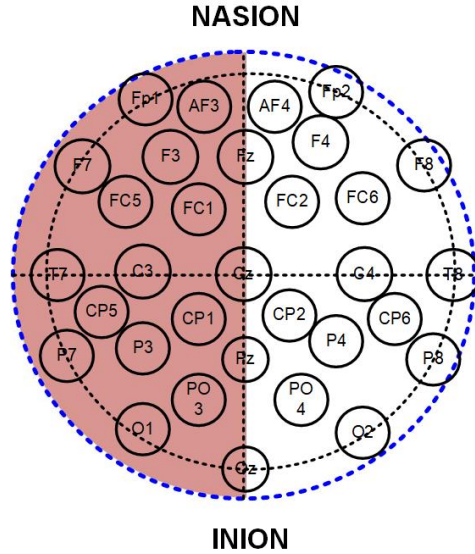
Figure 2-5: EEG 10-20 System Layout.

2.5   Emotion Prediction Data Sets and Labels

The data set selection is a very critical choice for the design and verification of an emotions classification system. A researcher may claim 100% accurate classification results on a self-collected data set of a few subjects. However, the system would fail to generalize on other data sets. Therefore, we have tried to identify the data sets which are being used by the top international scientific research publication forums related to biomedical systems and healthcare. DEAP and SEED are the two most popular and widely used data sets for emotions classification using physiological (EEG, ECG, EMG, EOG, etc.) signals [24], [36].

Table 2.2 summarizes the DEAP and SEED data sets. The DEAP data set provides the EEG recordings of 32 participants including 17 males and 15 females, respectively. The SEED data set provides the EEG recordings of 15 participants including 7 males and 8 females, respectively. The total number of EEG channels recorded in DEAP and SEED data sets are 32 and 64, respectively. The DEAP data set provides the emotion classification using valence and arousal labels scaled between 1 to 7. The SEED data set provides the emotion classification label using valence scaled as -1,0 and 1. The SEED and DEAP data sets were recorded on the sampling frequencies of 512 Hz and 1000 Hz, respectively.

The prediction of human emotion requires positive and negative valence or arousal labels

21

Table 2.2: Emotions Classification Data Sets Summary.

Data Set	Participants (#)	EEG Channels (#)	Labels	Classification Scale
DEAP	32 (17,15)	32	Valence, Arousal	1-7
SEED	15 (7,8)	64	Valence	-1,0,1

using DEAP and SEED data sets. The SEED data set already provides the positive and negative valence labels for each EEG signal. In the DEAP data set, the valence and arousal values are divided into positive and negative labels through low and high threshold values between 1 to 7. For example, the threshold value of 5 defines the valence and arousal values ≤ 5 as negative and > 5 as positive. Table 2.3 list the valence and arousal labels of a subject (Subject # 1) in the DEAP data set with a threshold value of 5 for the first ten emotional trials of the subject.

Table 2.3: Labels Summary for First 10 Emotional Trials in DEAP (Subject # 1).

Trial No	Valence Value	Arousal Value	Valence Label	Arousal Label
1	7.71	7.6	HIGH	HIGH
2	8.1	7.31	HIGH	HIGH
3	8.58	7.54	HIGH	HIGH
4	4.94	6.01	LOW	HIGH
5	6.96	3.92	HIGH	LOW
6	8.27	3.92	HIGH	LOW
7	7.44	3.73	HIGH	LOW
8	7.32	2.55	HIGH	LOW
9	4.04	3.29	LOW	LOW
10	1.99	4.86	LOW	LOW

2.6 Machine Learning for Emotions Prediction

ML based EEG emotions prediction systems targets extracting a minimum amount of useful and relevant information from a large raw EEG data. The minimum information derived from a large amount of data predicts human emotions. The ML process includes the calculation of suitable features and then applying an appropriate classification algorithm. The ML algorithm is either on-sensor by integrating an on-chip DBE processor or through cloud computing. A summary of both approaches is discussed below:-

- **Cloud Computing Based Emotions Prediction:** Cloud computing for emotional

prediction would require very high energy during computations and data transfer. These systems could not provide 24/7 assistance to the children as they are highly inappropriate for wearable systems. The primary reason for the unsuitability includes constant data transmission for cloud computing. These systems cannot be operated with small coin cell batteries and provide long battery timings. The continuous data transmission would also generate heat due to large energy consumption causing discomfort for the users. These systems would also cause issues related to the privacy and security of the patient's data.

- **Fully On-Sensor SoC Based Emotions Prediction:** The fully on-sensor SoC-based emotions prediction system using EEG signals requires integrating an AFE with a DBE processor. The AFE is responsible for acquiring low-noise EEG signals. The DBE processor is responsible for the calculation of the features and then applying the classification algorithm to the calculated features. The on-sensor and fully-on-chip emotions prediction system has several benefits as it avoids high communication costs, provides low latency, requires lower energy, and can provide long battery timings. The fully on-chip AFE and DBE implementation also avoids any data security and privacy issues.

This thesis targets the design and development of a DBE processor for emotion prediction. Therefore, we would discuss the design challenges for a fully on-chip DBE emotions prediction processor.

2.7 On-Chip DBE Processor Design Challenges

A ML process consists of two phases including "training" and "inference". The training phase includes learning a model using the available EEG data and the corresponding emotions. The learned model establishes a relation between the EEG signals and the corresponding emotions. The inference phase applies the learned model to predict the emotions of unseen EEG data. The training phase is carried out offline, whereas the inference is performed on-chip. The inference process consists of two parts including feature extraction and classification. The feature extraction process requires extracting the most suitable features to distinguish between positive and negative emotions. The features can be either

customized or DNN based. Some of the challenges to be considered before the selection of any ML or deep learning algorithm for a DBE processor are summarized below:-

- **Customized vs Deep Learning Based Feature Extraction:** The customized feature extraction process discovers the best and minimum number of features for a particular application to provide the maximum classification results. However, deep neural networks do not distinguish between feature extraction and classification. The deep learning-based approach directly maps the output labels to the input EEG data without any separate feature extraction process.

  The difference between conventional ML and deep learning-based feature extraction is depicted in Figure 2-6. In the conventional ML or "hand-crafted" feature extraction, the extracted features are forwarded to the classifier. The classifier utilizes these features to predict a negative or positive emotion. However, in deep learned features, the positive or negative labels of the emotion are directly calculated from the input EEG signals.

Machine Learning

EEG Signal

Feature Extraction → Classification → Neg. Emotion / Pos. Emotion

Deep Learning

EEG Signal

Feature Extraction + Classification → Neg. Emotion / Pos. Emotion
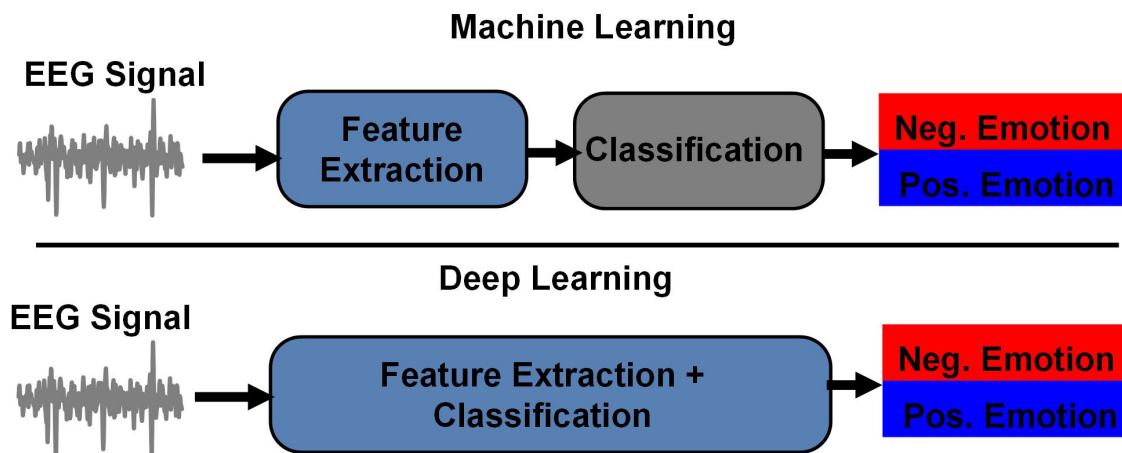
Figure 2-6: Customized Feature Extraction vs Deep Learning.

Deep-learned features can deliver high accuracy as compared to customized features. But, they require a large number of weights to directly map the EEG signals to the positive or negative emotions. The number of weights can be in millions and the required memory can be in gigabytes. The computational cost involved to deal with these weights is also very high, which would consume a very high amount of energy. Therefore, we have opted for the customized features and avoided deep-

learned features for our SoC-based emotion prediction DBE processors.

- **Feature Selection and Customization:** The ML classification tools and applications in MATLAB, python, etc. are primarily designed for "software-based systems". The classification results provided by these tools do not consider the hardware cost of the ML algorithm. Therefore, we have to analyze and calculate the hardware cost of selected features before their implementation. For example, the hardware cost of wavelet entropy is > 100X higher than the PSD of an EEG signal. Therefore, a hardware designer has to analyze the projected hardware cost including area, power, and energy efficiency before the selection of features. The features calculated using ML tools also have 64-bit double-precision values by default. However, the hardware designers have to analyze the trade-off between the classification accuracy and the hardware cost. The hardware cost directly depends on the number of bits. Therefore, we have to select the precision for the extracted features and classifier after an extensive number of bits vs classification results analysis.

- **Classifier Selection and Customization:** The ML algorithms utilize several classification algorithms including SVM, decision tree (DT), k nearest neighbor (KNN), naive Bayes (NB), or artificial neural networks (ANN). The hardware designers have to consider multiple parameters in classifier selection other than classification results. These parameters include kernel selection (Linear or Gaussian) in SVM, tree depth in DT, number of neighbors in KNN, and the number of nodes with the activation function selection (linear, relu, sigmoid, tanh, etc.) in DNN. The selection of each parameter has a significant impact on the hardware implementation cost. For example, the hardware implementation of the relu function is quite simple as compared to the sigmoid or tanh function.

2.8 Summary

In this chapter, I have provided the necessary background knowledge to the readers of this thesis. I have explained the following things:-

- What are human emotions and how are they measured?

- What are ML, classification, and regression?

- What is feature extraction and classification in ML?

- What are an EEG Signal, iEEG, and Scalp EEG?

- An introduction to the emotions prediction data sets and labels.

- An introduction to use ML for emotions prediction.

- The design challenges and constraints to design a DBE processor for emotions prediction.

Chapter 3

Literature Review

A significant amount of research has been done in the area of internal human emotion recognition of EEG signals due to the dependency of daily activities on emotions and technological advancements in miniaturized EEG acquisition systems [15]. Negative emotions are a major personality disorder, especially in ASD patients [37]. People with negative emotions disorder frequently encounter negative emotions [37]. Negative affectivity refers to a personality attribute to continuously encounter negative emotions irrespective of the situation [38]. ASD patients have a higher frequency of negative emotions and are strongly affected by negative affectivity and face negative emotion outbursts (NEOB) [39]. NEOB are unpredicted, sudden, and uncontrolled bursts of negative emotions leading to self-injuries and suicide attempts.

Despite the significant efforts done in the past two decades, still, no reliable wearable EEG-based system is available for NOEB, or negative emotions prediction [38],[39]. Moreover, even the focus of the research is more on a software-based solution for emotion detection, which has real-time limitations due to large decision latency. To ensure maximum benefit from emotion detection, it necessitates a wearable device like a hearing aid, smartwatch, or headband to capture real-time emotion and provide closed-loop feedback for early and meaningful intervention [40].

I conducted an extensive literature review on the previous works targeting emotion prediction using EEG signals. Since I was targeting a hardware-based SoC for emotional prediction. The inclusion criteria were based on the hardware implementation of the system, classification results, the data sets utilized for the system validation, the number of EEG

channels, feature and classifier selection, area, and power utilized by the system.

The literature review concluded that previous works targeting emotions classification using EEG signals can be categorized into four categories including software implementations with a local (private) data set, based on a large number of EEG channels, utilizing complex features, classifiers, and hardware implementations. All the previous studies categorize among the four categories discussed below:-

- **Category 1: Systems Validated Using Local Data Sets:** DEAP and SEED are the two most widely used data sets for emotion prediction and are considered a benchmark. These data sets contain a large number of trials for emotion prediction with a large number of subjects. These data sets contain subjects from both male and female genders, different age groups, and balanced distribution of positive and negative emotion labels. Some past works have claimed excellent classification results using local (private) data sets of a few subjects [41]. However, their algorithms are not validated using benchmark data sets, and hence not suitable for real-time implementations.

- **Category 2: Large Channel Count:** Some systems validated using public EEG data sets provide excellent classification results with $\geq 90\%$ accuracy [42]. However, these systems have utilized a large number of EEG channels. The researchers working on SoC-based systems for wearable applications have strongly recommended avoiding a channel count > 8 [27]. The higher number of channels is a bottleneck in the system's implementation and real-time performance due to area, power, and energy constraints. The patient's comfort is also compromised due to a large number of channels [27].

- **Category 3: Complex Features and Classification Algorithms:** The emotion prediction algorithms with fairly high classification results and low channel count are validated on benchmark data sets. However, the features and classification algorithms utilized by these systems are highly complex and unfeasible for wearable applications. The highly complex features and classification algorithms including stacked auto-encoders, dense convolutional, and deep neural networks, etc. have highly complex hardware realization [43],[44]. The hardware implementations of these features and classification algorithms would be expensive (more chip area) and have low battery timings.

30

- *Category 4: Hardware Implementations:* Some systems have provided hardware-based solutions for emotion prediction, and have achieved very good ($\geq 75\%$) classification results. However, these systems have either provided only FPGA-based implementations [44], or lack fully integrated SoC-based solutions [26]. Their hardware implementations also have high area and power and hence could not provide long battery timings. These systems are also not validated using multiple benchmark data sets.

.

3.1 Previous State-of-the-Art Works

An organized review of previous studies that have aimed to provide emotional prediction using EEG signals is presented in Table 4.1. The hardware implementation, benchmark or public data set validation, the number of EEG channels (Channel #), list of features, classification algorithm (CLF), classification results (Accuracy %), area and classification energy of the system (for hardware implementations) of the previous works are listed in the table. I performed an extensive literature review of the previous state-of-the-art works including these works. The SoC-based processors in this Ph.D. project were designed after keeping in view these performance constraints, and previous works.

Panagiotis [41] proposed a software-based emotions classification algorithm using only three EEG channels. They utilized the higher order crossing (HOC) features. Multiple-order polynomial kernel-based SVM classifiers were used for the classification. They achieved excellent classification results (83.3% accuracy). The EEG channels used by them correspond to the frontal-parietal locations. These locations would cause minimum discomfort for ASD patients. However, they validated their algorithm on a private data set of limited (15) subjects.

T. Song [45] proposed a software-based emotions classification algorithm using 62 EEG channels. They utilized the PSD, HOC, Hjorth (HJ), Hilbert-Huang transform (HHT), and short-term Fourier transform (STFT) features. Three attention long short-term memory (A-LSTM) architecture was used for the classification. The final class is decided by the majority voting. The A-LSTM architecture is highly unfeasible for SoC implementations due to huge memory requirements. A large number of channels is another limitation of this

Table 3.1: Literature Review.

Ref	Hardware (Yes/No)	Data Set (Pub/Pvt)	Channel #	Features	CLF	Acc (%)	Area/Power) (mm/uJ)
[41]	No	Private	3	HOC	SVM(Poly)	83.3	—
[45]	No	Private	62	PSD,HOC, HJ, HHT, STFT	A-LSTM	71.8	—
[46]	No	Public	15	3DCNN	3DCNN	97.3	—
[24]	No	Public	16	PSD	NB	62	—
[42]	No	Public	62	PSD,DE, DASM RASM, DCAU	GELM	91.1	—
[36]	No	Public	62	DE	DBN	86.1	—
[47]	No	Public	1	AlexNet	SVM (RBF)	90.2	—
[48]	No	Public	32	SK, KT,HJ, FD, WE, FC.	SVM (RBF)	88.5	—
[49]	No	Public	32	PSD	LSTM-RNN	79.5	—
[50]	No	Public	32	DE	CNN	90.6	—
[51]	No	Public	62	DE	CNN	78	—
[44]	No	Public	32	Hybrid CNN	Hybrid CNN	80.8	—
[43]	No	Public	32	STF,RASP	LSTM	60.2	—
[25]	Yes	Public	14	PSD (Welch)	CNN	83.1	41,409*/150
[26]	Yes	Public	6	Entropy, STFT, DASM	CNN	83.37	3.35/76.61
[52]	Yes	Public	6	ASR, STFT, DASM	CNN	83.7	–/29.51
[53]	Yes	Private	8	LRCNN	LRCNN	88.34	1.28/48.24
[54]	Yes	Public	12	LRCNN	RCNN	82.88	—

work. They also validated their system on a local data set with 23 participants. The private data set validation is another issue in their system.

Shuaiqi [46] proposed a three-dimensional convolutional neural network (3DCNN) based algorithm for emotion prediction. They utilized 15 EEG channels for emotion prediction and achieved excellent classification results (97.3% accuracy). The feature extraction was also based on the proposed 3DCNN. They validated their system on the SEED data set. The number of EEG channels and the 3DCNN architecture for feature extraction and classification algorithms require a large amount of memory and consume high power. Hence, they are not suitable for SoC systems with strict battery timing constraints.

Koelstra [24] proposed an emotions classification algorithm using 16 EEG channels. They utilized the PSD of different frequency bands as features and the NB classifier as the classification algorithms and achieved 62% classification accuracy. The algorithm was validated on the DEAP data set. The main limitations of this work are low classification results, lack of multiple data sets validation, and a large number of EEG channels.

Peiyang [42] proposed an emotions classification algorithm using PSD, differential entropy (DE), differential asymmetry (DASM), rational asymmetry (RASM), asymmetry (ASM), and differential caudality (DCAU) features. They utilized 64 EEG channels and achieved 91.1% classification results. A customized Graph regularized Extreme Learning Machine (GELM) classification algorithm was utilized in this work. Although they achieved excellent classification results using multiple benchmark data sets, their algorithm also utilizes a large number of channels. The GELM algorithm was primarily designed for software-based implementations and was not suitable for hardware implementation. The DE, RASM, ASM, and DCAU features would also have highly complex hardware realization.

Demir [47] proposed an emotions classification algorithm using a single EEG channel and achieved excellent classification results using the DEAP data set. They utilized the AlexNet CNN for the feature extraction and radial basis kernel function (RBF) SVM for classification. The Alexnet CNN has a highly complex architecture and requires millions of classification parameters (weights and biases) for hardware realization.

Soleymani [49] proposed an emotions classification algorithm using a local data set of ten participants. They used 32 EEG channels for the classification. They used the PSD of different frequency bands as features and Long Short Term Memory Recurrent Neural Network (LSTM-RNN) architecture as the classification algorithm. The main limitations

of this work were the large number (32) of EEG channels and a private data set validation. The LSTM-RNN architecture is also highly complex for hardware realization.

Zhongke [50] proposed an emotions classification algorithm using DEAP and SEED data sets. They used 32 EEG channels for the classification. They used DE as the feature and a CNN for the classification algorithm and achieved excellent (90.6%) classification results.

Luo [51] proposed an emotions classification algorithm using multiple public EEG data sets. They achieved 78% classification results using 64 EEG channels for the DEAP and SEED data sets respectively for emotions classification. The DE and CNN were used for the feature extraction and classification respectively. The main limitations of this work were the large number of channels and complex hardware realization for DE and CNN.

Hector [25] implemented an FPGA-based hardware algorithm for emotions classification using publicly available EEG data sets. They utilized 14 EEG channels and provided 83.1% classification results. They used PSD and CNN for the features extraction and classification algorithm respectively. Although they provided hardware (FPGA) implementation, they did not provide SoC implementation. They utilized 41,409 look-up tables and flip flops using a Spartan-6 FPGA board with a power consumption of 150 mW. The main limitations of this work are a large number of EEG channels, lack of SoC implementation, and high area and power.

Fang [26] designed and developed an SoC using a 28-nm CMOS process for emotion prediction. They utilized only 6 EEG channels and achieved excellent(83.37%) classification results using public EEG data sets. The main limitations of this work are the lack of a fully-integrated SoC solution. They did not provide SoC implementation for the feature extraction engine. The features are extracted offline and fused in the form of images. The feature images are then transmitted to the SoC using blue-tooth communication and a Spartan 3E FPGA. The area and power consumption of this system were $3.35mm^2$ and 76.61 mW respectively.

Yang [53] proposed an SoC-based hardware solution using a 16nm FET process. They utilized a private data set of 52 subjects for the validation of their system. They utilized a long-term recurrent neural network (LRCN) for feature extraction and classification. The area and power consumption of the system are $1.29mm^2$ and 48.24mW, respectively. They also did not provide a fully integrated SoC solution and the feature extraction process was carried out offline (outside SoC). They also lack public data set validation which limits their

robustness.

K. Wang [54] proposed an emotions prediction network-on-chip using DEAP data set with 83% accuracy. They utilized a CNN which was implemented on a 28 nm CMOS process. They utilized an EEG channel subset of 12 channels for the emotions prediction. The main frailty of their work was a large number of channels, lack of multiple BM-DS validations, and offline feature extraction. Since they targeted a network-on-chip, they only implemented the CNN on the chip. The offline feature extraction forwards a set of windowed images after noise removal, BPF, STFT, channel reconstruction, and image generation. They also did not report the area, power, and classification energy of their CNN-based network-on-chip implementation.

C. Ardito [55] proposed an emotions prediction algorithm using five EEG channels. They tested and validated their algorithm using the DREAMER data set. They utilized a single-dimensional CNN in their emotions prediction algorithm, Although they achieved excellent classification results (93%), they also lack multiple BM-DS validations and hardware implementation.

X. Shen [56] proposed a contrast learning method for inter-subject alignment, which proposed an emotions prediction algorithm using SEED and a private data set THU-EP. The main novelty of their work was cross-subject or subject-independent emotion recognition. Their contrastive learning algorithm utilized a CNN and provided excellent (86.4%) classification results. But they lack multiple BM-DS validations, utilize a large number of channels, and lack hardware implementation.

J. Li [57] proposed a single-channel emotion recognition algorithm with excellent classification (82.18%) results. They validated their algorithm on multiple BM-DS including DEAP, SEED, and MAHNOOB [57]. They utilized subject-dependent variable channels for variable time instances in their algorithm, which is not suitable for a wearable SoC design.

A. Menon [58] designed and proposed an energy-efficient hyper-dimensional computing processor. The processor was implemented using a 28 nm CMOS process and validated using a non-frequently used emotions recognition data set AMIGOS.

3.2   Summary

The review and analysis of the previous works provided me with the targets of multiple public data sets validation, a lower number of EEG channels, high classification accuracy, and a fully integrated SoC solution. Therefore, I designed and developed two SoC-based DBE processors (DBE^1 and DBE^2) using TSMC 180 nm CMOS process. The proposed DBE processors are discussed later in Chapter 4 and Chapter 5 in detail [38], [39]. The channel count, classification accuracy, public or private data set validation, and the hardware or software-based implementation of the previous state-of-the-art works is depicted in Figure 3-1.
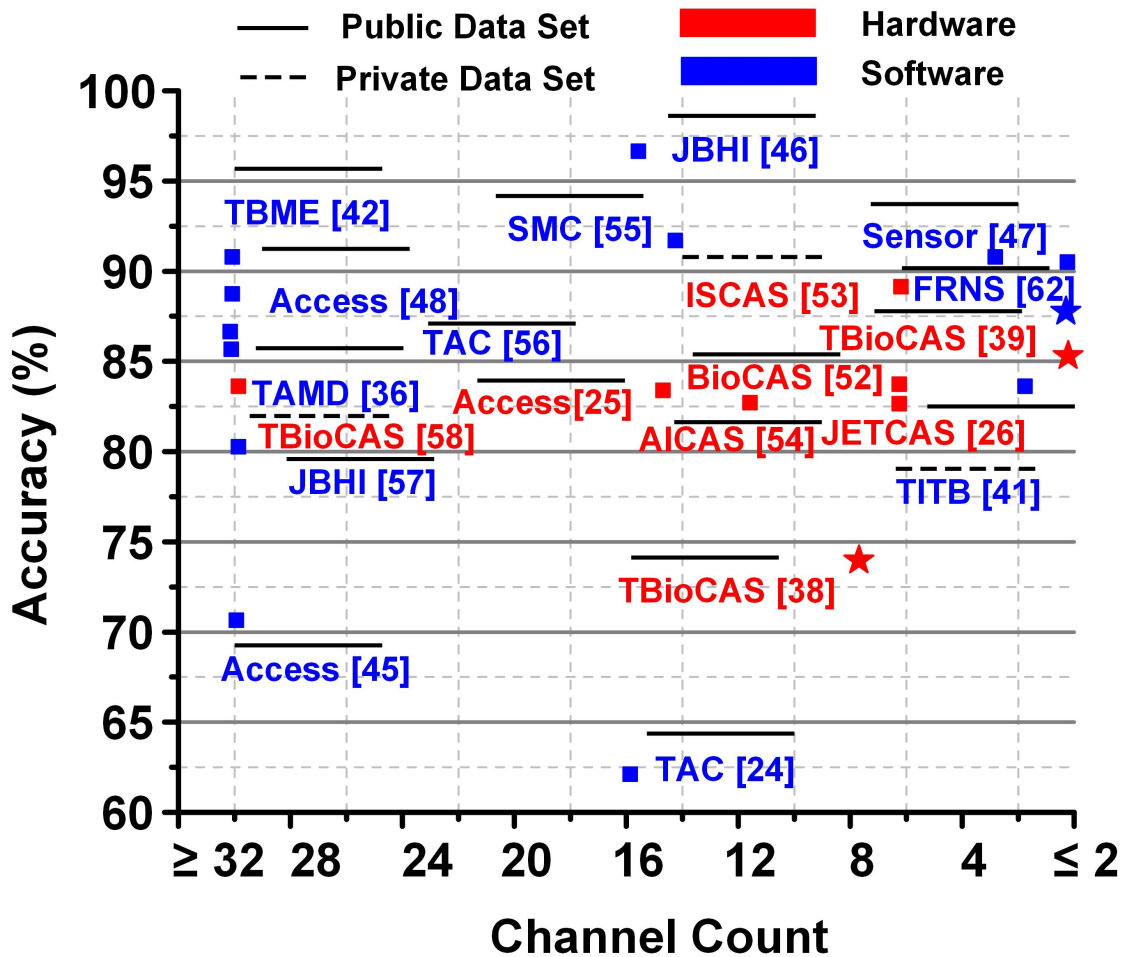
Figure 3-1: Previous State-of-the-Art Works.

It can be observed that the majority of the previous works utilized ≥ 16 EEG channels.

Some of the previous works provided private data sets validation. Some of the previous works have provided low classification results. Therefore, I targeted a high classification accuracy ($\geq 80\%$) with a minimum number of channels (≤ 8), and multiple public data sets validation. My first SoC-based DBE processor (DBE^1) predicted human emotions with 73.4% classification accuracy and utilized only eight EEG channels [39]. This 180 nm CMOS processor used PSD and linear SVM for the feature extraction and classification, respectively. The processor utilized an area and power of $5.4mm^2$ and 16 μJ, respectively. The processor was validated on multiple public data sets. The second SoC-based DBE processor (DBE^2) performed emotions prediction with 85.4% accuracy using only two EEG channels. The processor used zero crossings and skewness for the feature extraction, and a feed-forward deep neural network for the classification algorithm. The processor utilized an area and power of $16mm^2$ and 10.13 μJ, respectively. The processor was also validated using multiple public data sets.

Chapter 4

DBE^1: 1st Generation Digital Back-End Processor for Negative Emotions Prediction

The design of the 1st generation DBE processor (DBE^1) is presented in this chapter. The challenges related to the processor design for the suitable channels, features, classifier identification, and hardware implementation are discussed. The classification algorithm, hardware implementation with optimization of the identified features, and the classification algorithm to minimize area, and classification energy of the proposed DBE^1 are explained in this chapter. Section 4.1 explains the process of emotion prediction using EEG signals. The choice of data sets, labels for emotion prediction, and the challenges in emotion prediction using EEG signals is explained in this section. Section 4.2 explains the ML algorithm used for negative emotions prediction in DBE^1. The selection of a suitable subset of EEG channels, features, and classification algorithm selection are explained in this section. Section 4.3 explains the hardware architecture for the SoC implementation of the selected ML algorithm in DBE^1. The overall architecture of DBE^1, hardware implementation of EEG pre-processing unit, feature extraction engine including feature normalization unit, classification unit, and the parameters register for classification parameters upload is explained in this section. The classification performance of DBE^1 including classification results, chip performance summary, and performance comparison with the other state-of-the-art

works are discussed in section 4.4. Section 4.5 summarizes the overall achievements and shortcomings of DBE^1.

4.1 EEG Emotions Prediction

Emotion prediction using EEG signals is a very challenging problem due to multiple factors including limited benchmark data sets, unavailability of extensive analysis for suitable features with channels identification, and subject-dependent emotional patterns [59]. These challenges are a major bottleneck for the robustness and reliability of emotion prediction systems [59]. Therefore, emotions prediction is considered a very challenging task in comparison to other EEG classification tasks like epilepsy prediction or motor imagery classification [60]-[61]. Another major challenge for the EEG emotions prediction is the unavailability of medically verified annotated emotions and emotion triggers [62]. The epilepsy prediction data sets for example contain the medically verified and annotated EEG patterns during and before seizures [27]. The labels for emotion prediction are marked on the user experience and feelings, which may not depict the actual state of the mind resulting in intermixed emotions [63]. That's why emotion prediction is considered a testing task and many state-of-the-art works have reported lower classification results [38].

4.1.1 Emotions Prediction Scale and Data Sets

Human emotions are very complex and are a combination of different neurological, behavioral, and psychological processes [64]. The human emotions are mapped either in a discrete number of emotions or a bi-directional valence-arousal scale as explained earlier in Chapter 2 (Background Knowledge) [65]. I have utilized Russel's 2-D valence-arousal scale for the emotions prediction in this DBE^1 processor for negative emotions prediction [34].

The DEAP and SEED data sets for emotions prediction using EEG and other physiological signals were utilized for the design and validation of the DBE^1 processor [38]. The valence and arousal scales in the DEAP data set and the valence scale in the SEED data set are used for emotion prediction. The DEAP data set has mapped the valence and arousal of each participant from 1 (minimum) to 9 (maximum) [24]. The positive and negative emotions classification in the DEAP data set is shown in Figure 4-1.

The valence and arousal values mapped between $1(T_{min})$ to $9(T_{max})$ were labeled as

40
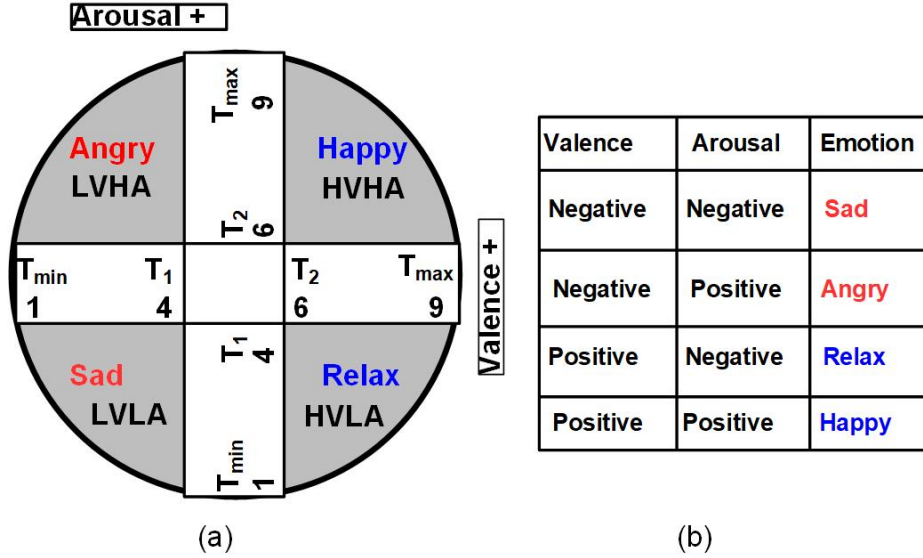
Figure 4-1: (a) Valence & Arousal Scale for DEAP (b) List of Predicted Emotions in DEAP.

positive or negative using the threshold values of $4(T_1)$ and $6(T_2)$. The valence and arousal values between $1(T_{min})$ to $4(T_1)$ were labeled as negative. The valence and arousal values between $6(T_2)$ to $9(T_{max})$ were labeled as positive as shown in Figure 4-1 (a). The negative (low) valence with negative (low) arousal, negative (low) valence with positive (high) arousal, positive (high) valence with negative (low) arousal, and positive (high) valence with positive (high) arousal are mapped as sad, angry, relaxed, and happy emotions, respectively as shown in Figure 4-1 (b). The SEED data set provides valence labels only as negative, neutral, or positive [36]. The positive or negative valence prediction was performed for DBE^1 processor and the neutral labels were discarded.

4.1.2 Emotions Prediction Challenges

The emotion classification is a very daunting challenge due to the unavailability of visible markers to distinguish between different emotions [59]. The time domain EEG signals for different time segments and frequency response in different frequency bands are used for emotions classification using EEG signals [59].

The analysis of time-domain EEG signals and the frequency domain response of different frequency bands are mostly used for different physiological signal classification problems. I analyzed different time segments of EEG signals along with their frequency responses and

the band-passed signals in different frequency bands.

Figure 4-2 shows the time-domain plot of Subject # 01 in the DEAP data set for F7 channel. Figure 4-2 (a) shows the EEG signal for emotional trial # 11 with valence and arousal values of 2.99 and 2.36, respectively. The valence and arousal values of 2.99 and 2.36 correspond to low valence and low arousal (LVLA) with the selected threshold values. Figure 4-2 (b) shows the EEG signal for emotional trial # 15 with valence and arousal values of 3.17 and 8.08, respectively. The valence and arousal values of 3.17 and 8.08 correspond to low valence and high arousal (LVHA) with the selected threshold values. Figure 4-2 (c) shows the EEG signal for emotional trial #22 with the valence and arousal values of 7.09 and 2.08, respectively. The valence and arousal values of 7.09 and 2.08 correspond to high valence and low arousal (HVLA) with the selected threshold values. Figure 4-2 (d) shows the EEG signal for emotional trial #25 with the valence and arousal values of 8.86 and 7.21, respectively. The valence and arousal values of 8.86 and 7.21 correspond to high valence and high arousal (HVHA) with the selected threshold values.

The extensive analysis of different time domain EEG segments for LVLA, LVHA, HVLA, or HVHA does not show any visible markers to differentiate between low and high, valence or arousal labels. The EEG signals were then band-passed from different frequency bands including theta, alpha, beta, and gamma band. Figure 4-3 shows the selected 15-second segment of the band-passed EEG signal of F7 channel in alpha, beta, and gamma bands for emotional trial numbers 11 (LVLA), 15 (LVHA), 22 (HVLA) and 25 (HVHA) in subject #01 from DEAP data set. The naked-eye observations of the band-passed signals also did not show any visible information to differentiate between low and high valence or arousal labels.

Similarly, Figure 4-4 shows the selected 15-second segment of the frequency response of the EEG signal of F7 channel in alpha (4-8 Hz), beta (12-30 Hz), and gamma (32-40 Hz) bands for emotional trial numbers 11 (LVLA), 15 (LVHA), 22 (HVLA), and 25 (HVHA) in subject #01 from the DEAP data set. The frequency response of the EEG signals also did not show any visible markers to predict the low or high valence and arousal labels using EEG signals. Therefore, complex ML or deep learning algorithms are required to extract the deeply unobserved information for emotion prediction using EEG signal [59].

The majority of the previous works in EEG-based emotion classification are on an algorithmic level (software-based) as discussed previously in Chapter 3. The software-based
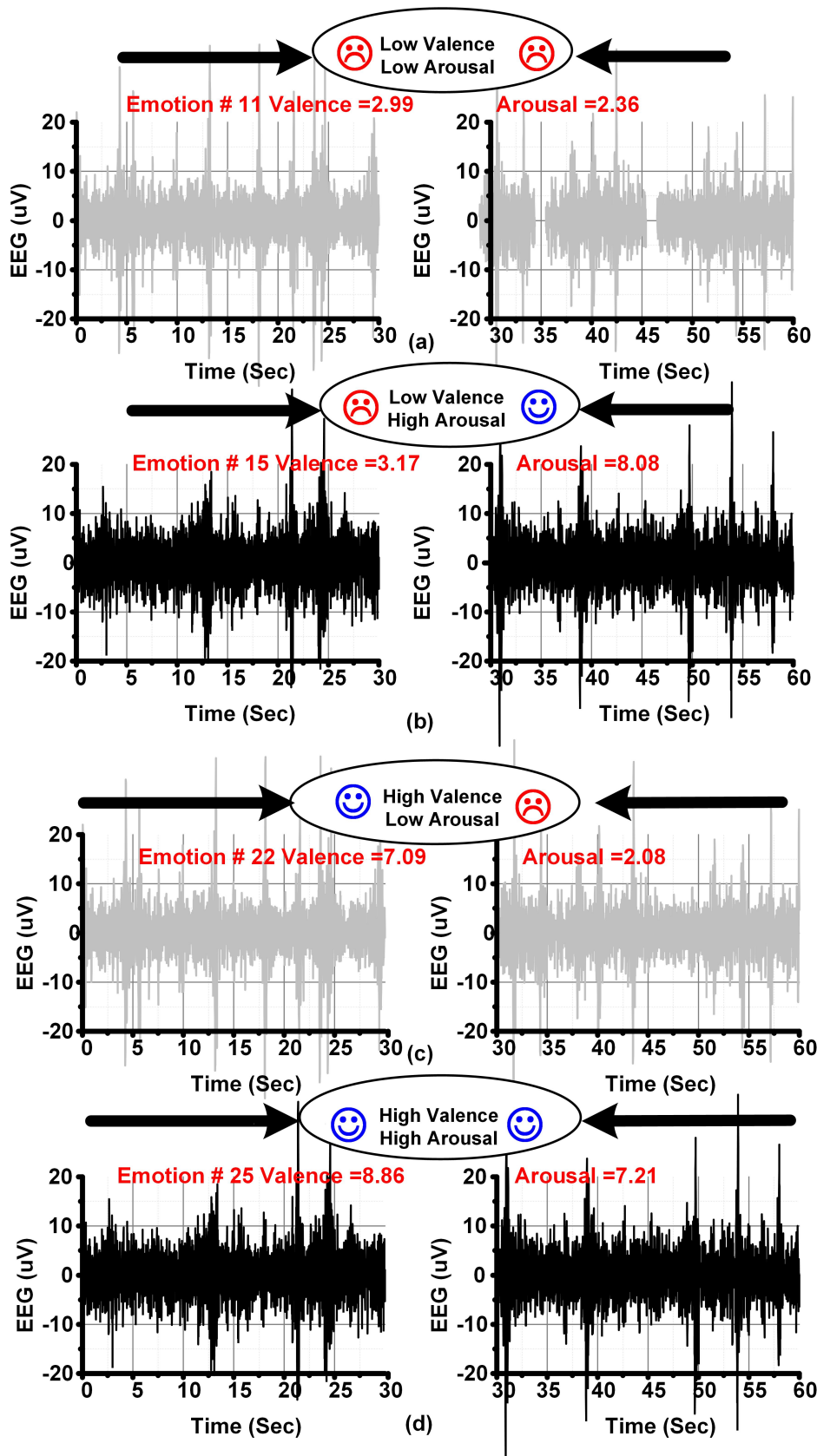
Figure 4-2: Time Domain Plot for (a) Low Valence Low Arousal (b) Low Valence High Arousal (c) High Valence Low Arousal and (d) High Valence High Arousal.
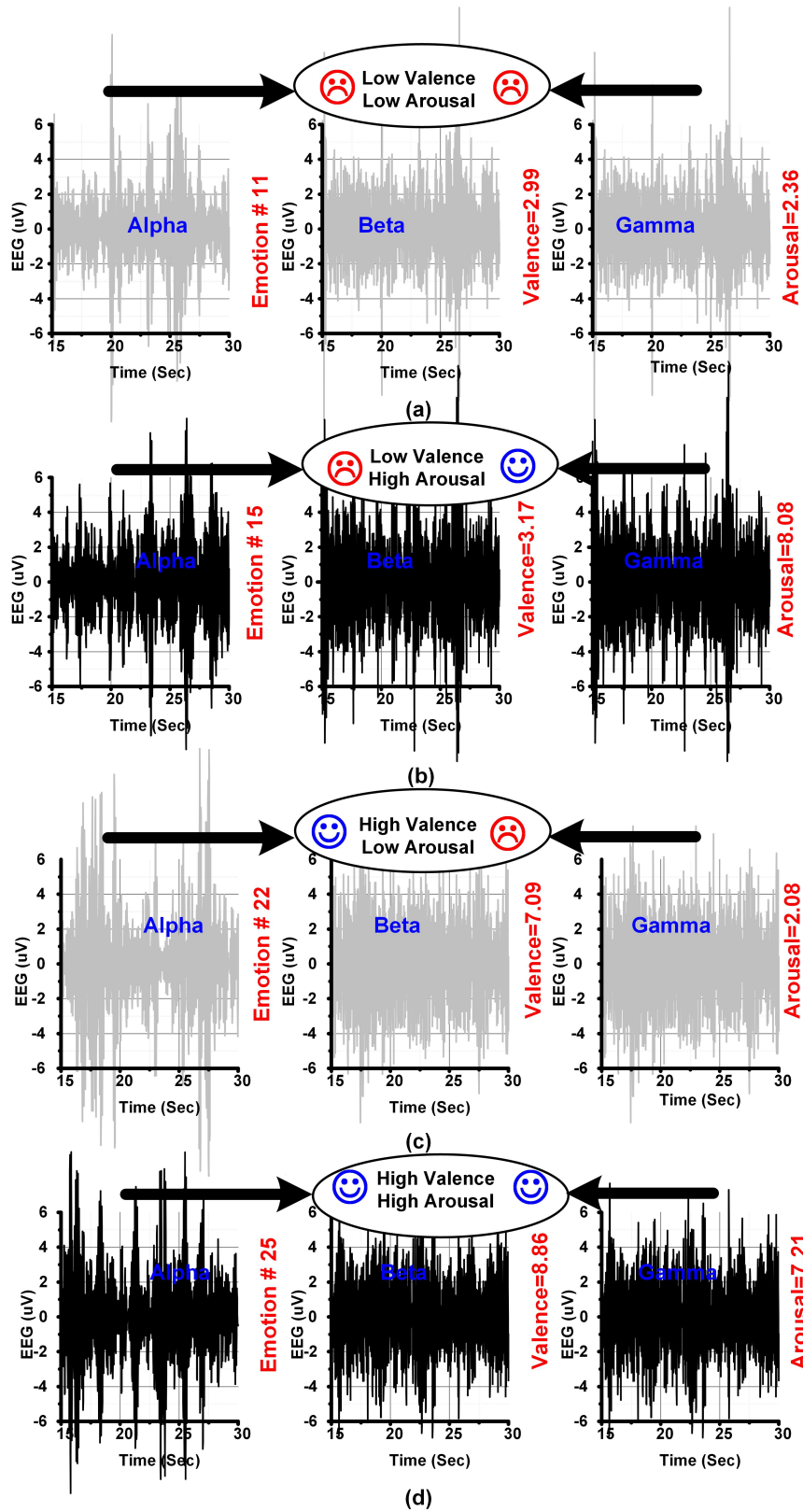
Figure 4-3: EEG plots in Alpha, Beta, and Gamma Bands for (a) Low Valence Low Arousal (b) Low Valence High Arousal (c) High Valence Low Arousal and (d)High Valence High Arousal.
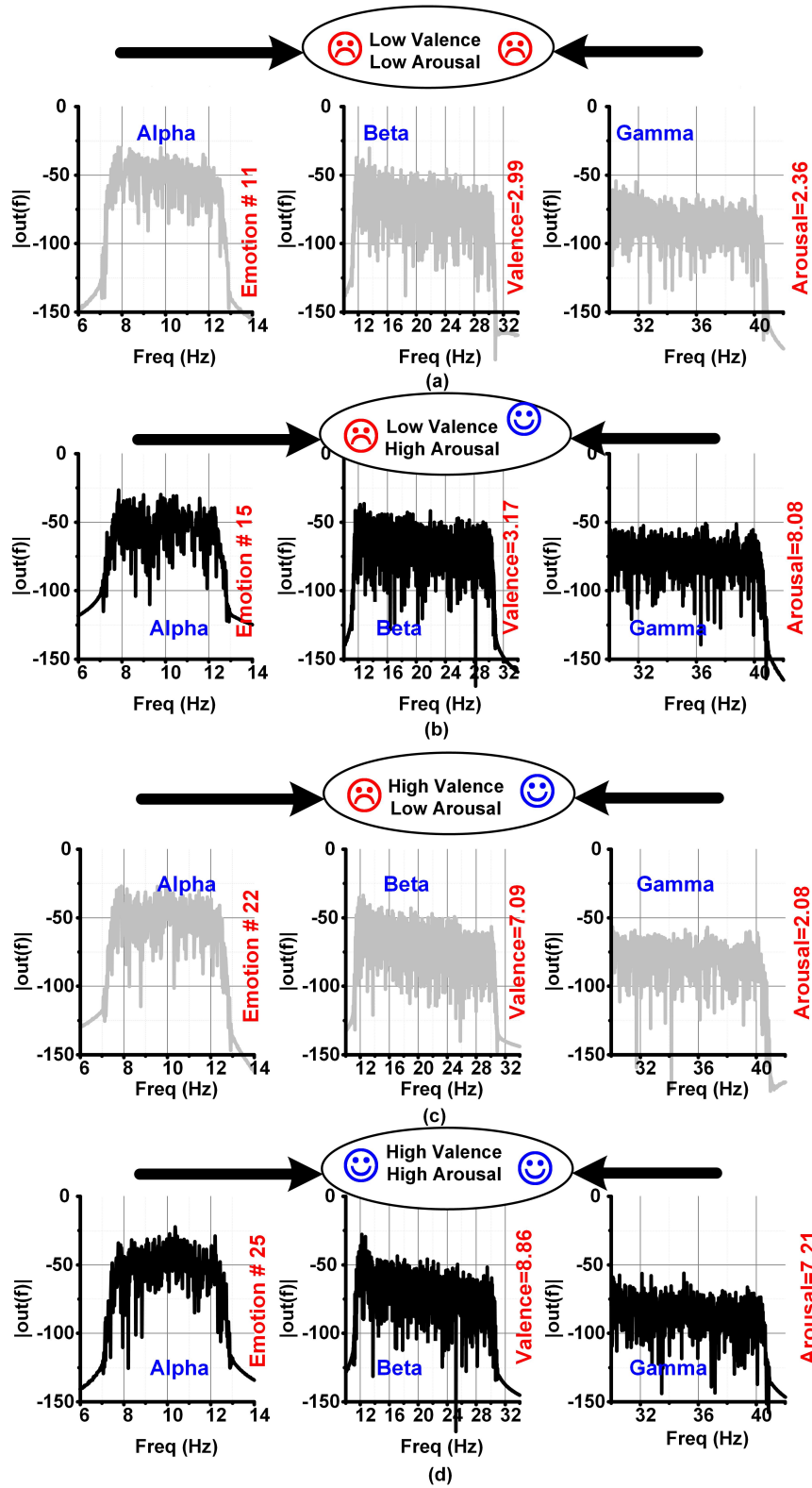
Figure 4-4: EEG Frequency Response in Alpha, Beta, and Gamma Bands for (a) Low Valence Low Arousal (b) Low Valence High Arousal (c) High Valence Low Arousal and (d)High Valence High Arousal.

algorithms have much flexibility in terms of the number of EEG channels, the number, and complexity of features, and the complexity of classification algorithm [66]. The previous works have utilized the maximum (64) number of EEG channels, a large number of complex features (HOC, HJ, HHT, STFT), or complex neural networks for emotions prediction [46],[45]. Since the primary target of these algorithms is to achieve the maximum classification results, they can utilize the maximum EEG channels along with large complex feature vectors and complex classification algorithms. However, a wearable SoC-based DBE processor has stringent constraints for the area, classification energy, and channel count [26], [38], [39]. Therefore, a hardware realizable feature vector utilizing limited EEG channels with a hardware feasible classification algorithm is required for our negative emotions prediction DBE processor's algorithm [66].

4.2 Negative Emotions Prediction Processor Algorithm

4.2.1 EEG Channels Selection

The channel selection is the process of choosing a subset of EEG channels suitable for emotions classification from the complete set of 32 or 64 channels provided in the data set [59]. The channel selection is a key procedure for the on-sensor integration of the proposed processor, which requires a limited number (≤ 16) of channels [67]. The on-sensor integration is required for efficient hardware implementation [67]. It is targeted to provide a wearable system for real-time negative emotion prediction [38], [39]. However, reducing the number of channels can lower the classification accuracy due to exclusion of useful information [66]. Therefore, it is important to analyze the effect of each channel on the classification results [66].

Channel selection involves the selection of a minimum number of channels without comprising the classification results. I therefore initially investigated the complete set of 32 EEG channels in the DEAP data set with different features. The complete set of 32 channels provided the best-case classification results of 54% accuracy. The complete set of 32 EEG channels is highly unfeasible for a wearable SoC implementation, in addition to the low classification results. The number of EEG channels being utilized for a wearable SoC system is a key to a feasible hardware implementation [26]. The reduction in the number of EEG channels directly affects the hardware complexity. Therefore, it is desired to utilize

the minimum number of EEG channels.

The channel reduction process requires finding the minimum number ($1 < N \leq 32$) of channels from the total set of 32 or 64 channels [24], [68]. The large channel count is not feasible due to high area, classification energy, and patient discomfort for wearable SoC systems [67]. I, therefore, targeted to utilize ≤ 16 EEG channels. Reducing the number of EEG channels directly lowers the hardware complexity and patient's discomfort [66]. Therefore, different combinations of 1, 2, 4, 8, and 16 EEG channels were investigated for negative emotion prediction. The frontal and temporal scalp locations are more suitable for the emotions prediction, based on the classification results and the patient's comfort [69]. A subset of 16 channels (FP1, AF3, F3, F7, FC5, FC1, T7, FP2, AF4, Fz, F4, F8, FC6, FC2, T8, and C4) including 15 frontal and temporal channels provided the best classification results among different 16 channel subsets.

The 16-channel subset (FP1, AF3, F3, F7, FC5, FC1, T7, FP2, AF4, Fz, F4, F8, FC6, FC2, T8, and C4) with the best case classification results was used to further reduce the channel count. Different subsets of 1, 2, 4, and 8 channels were utilized from these 16 channels. An eight-channel pool (F3, F4, F7, F8, AF3, AF4, T7, and T8) of electrodes provided the maximum classification results as compared to other 1,2,4 and 8-channel combinations. The classification results of the selected combinations of EEG channels, scalp placement locations of these channels, and their locations on the 10-20 placement system are shown in Figure 4-5. The temporal and forehead locations of EEG electrodes are more feasible for the long-term monitoring of the EEG data for ASD patients [70]. They are also less frustrating for continuous wearable systems and are strongly linked with human emotions and behaviors [59].

4.2.2 Feature Selection

The complexity of the FE algorithm would directly impact the overall area and energy of the negative emotions prediction DBE processor [38],[39]. The software-based systems on the other hand do not have any such limitations. Therefore, it is critical to analyze the hardware implementation cost of the features in addition to the classification results [38],[39]. The EEG signal of one-minute duration was used as a feature vector initially. Different ML classification algorithms were applied for the negative emotions prediction. The best-case classification algorithm obtained with this feature vector was $\approx 75\%$. But
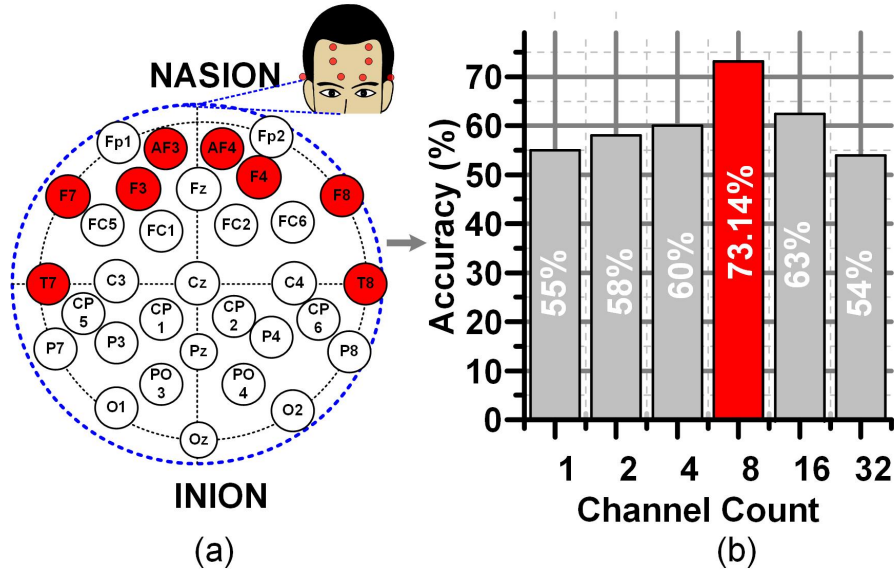
Figure 4-5: Scalp Locations of Selected Electrodes & Classification Results.

this feature vector is unrealizable from a hardware designer's perspective. It would require huge on-chip memory ($> 1MB$) for the FLP weights to be uploaded on-chip for the negative emotions prediction.

The number of features extracted per EEG channel also affects the hardware cost. I decided to limit it to four features per EEG channel. A set of different features including mean, root mean square, standard deviation, variance, Kurtosis, Hjorth parameters, entropy, and PSD in different frequency bands was experimented. The estimated hardware cost including area (number of gates) and the power consumption was calculated using TSMC 180 nm CMOS process. Table 4.1 lists some of the experimented features with the highest importance in different feature combinations and the classification accuracy for negative emotions prediction with a synthesized area (number of gates) and power using 180 nm CMOS process.

Common spatial patterns (CSP), wavelet transforms and PSD features provided maximum classification results for negative emotions prediction. The hardware implementation of the CSP requires covariance matrices with huge ($> 5MB$) memory requirements. Therefore, this feature vector was not selected. A feature vector containing PSD, AIHPD, and IHPR of EEG signals in the beta band (12 Hz to 30 Hz) was selected as shown in equation

48

Table 4.1: Features and Estimated Hardware Resources.

Feature	CLS Accuracy	Area* (# of Gates)	Power* (uJ)
Mean	55%	0.49k	10
Root Mean Square	50%	1.05k	30.3
Standard Deviation	60%	1.12k	22.2
Variance	45%	2.68k	50.04
Common Spatial Pattern	55%	11.75k	185
Hjorth Parameters	75.3%	8.06k	150.1
Entropy	68.3%	4.25k	78.3
PSD, AIHPD, IHPR	**71.3%**	**1.71k**	**34.2**
PSD, AIHPD, LIHPR	**73.4%**	**1.37k**	**20.1**

4.1.

The PSD is the spectral power of an EEG signal in a specific frequency band. The beta band ranging from 12 Hz to 30 Hz was used for our DBE^1 negative emotions prediction processor. AIHPD is the absolute difference of the PSD for the selected channels on the left(PSD_{Left}) and right (PSD_{Right}) side hemispheres. IHPR is the ratio of the PSD among selected channels on the left(PSD_{Left}) and right (PSD_{Right}) side hemispheres. PSD_{Left} is the PSD of the left side hemisphere EEG channels. PSD_{Right} is the PSD of the right side hemisphere channels. Equations 4.2, 4.3, 4.4, 4.5, and 4.6 are used to calculate the PSD, AIHPD, IHPR, PSD_{Left}, and PSD_{Right}, respectively.

$$FV = \left[\{PSD, IHPR, LIHPR\} \right] \tag{4.1}$$

$$PSD = \left[\sum_{i=0}^{N} [X_n] \right] \tag{4.2}$$

$$AIHPD = \left[|PSD_{Left} - PSD_{Right}| \right] \tag{4.3}$$

$$IHPR = \left[\frac{PSD_{Left}}{PSD_{Right}} \right] \tag{4.4}$$

$$PSD_{Left} = \left[PSD\{T7, F7, F3, AF3\} \right] \tag{4.5}$$

$$PSD_{Right} = \left[PSD\{T8, F8, F4, AF4\} \right] \tag{4.6}$$

A floating point (FLP) divider is required for the implementation and realization of equation 4.4. The FLP divider unit consumes high area and power, due to which division operations are avoided in hardware applications [71]. Therefore, to avoid the FLP division operation, and design a hardware feasible solution, it is required to approximate the IHPR feature without compromising the classification results.

IHPR was replaced by SIHPR initially [72]. SIHPR is a scaled ratio between PSD_{Left} and PSD_{Right} to avoid IHPR values lesser than 1 as shown in equation 4.7. The value of k was selected in way to ensure $SIHPR \geq 1$. The SIHPR provided good classification results (63% accuracy). However, this approximation of IHPR did not validate on multiple data sets including SEED data set [36].

$$SIHPR = \left[\frac{PSD_{Left}}{PSD_{Right}} * k \right] \tag{4.7}$$

Equation 4.4 for the IHPR feature can be rewritten as 4.8 and 4.9 using the difference conversion property of logarithms [73]. Equation 4.9 can be further simplified as 4.10 and 4.11. The LIHPR is the logarithm of the IHPR as shown in equation 4.12. The ratio-to-difference conversion properly of the logarithms is beneficial for the hard-

ware implementation of LIHPR using look-up-table (LUT) implementation. The hardware implementation of $log_2(PSD_{Left} \div PSD_{Right})$ for equation 4.8 requires the storage of all possible combinations of $log_2(PSD_{Left} \div PSD_{Right})$. However, the hardware implementation of $log_2(PSD_{Left}) - log_2(PSD_{Right})$ for equation 4.9 only requires the storage of $log_2(PSD_{Left})$ or $log_2(PSD_{Right})$, if both (PSD_{Left}) and (PSD_{Right}) have a similar range. Therefore, the calculation of LIHPR using the ratio-to-difference conversion property of logarithms requires only a square root of the memory required without using this property.

$$IHPR = \left[2^{log2(\frac{PSD_{Left}}{PSD_{Right}})} \right] \tag{4.8}$$

$$IHPR = \left[2^{log2(PSD_{Left} - log_2 PSD_{Right})} \right] \tag{4.9}$$

$$IHPR = \left[2^{log2(\frac{PSD_{Left}}{PSD_{Right}})} \right] \tag{4.10}$$

$$IHPR = \left[2^{LIHPR} \right] \tag{4.11}$$

$$LIHPR = \left[log_2(\frac{PSD_{Left}}{PSD_{Right}}) \right] \tag{4.12}$$

The IHPR feature requires the calculation of the power of 2 ($2^{\wedge}$) or anti-log base 2 of the IHPR as shown in equation 4.11. Therefore, the original IHPR feature was approximated with the LIHPR. The IHPR feature is embedded inside the LIHPR feature, which can be retrieved by the anti-log base 2 operations. Moreover, the LIHPR feature does not hamper the classification results in comparison to IHPR.

The IHPR and LIHPR of a subject (subject # 27) in the DEAP data set for two EEG channels (F7, F8) is shown in Figure 4-6. Figure 4-6 (a) and Figure 4-6 (b) show the

increasing trend of IHPR and LIHPR, respectively. The IHPR and LIHPR for a constant PSD_{Right} and variable PSD_{Left} values in ascending order are shown in Figure 4-6 (a) and Figure 4-6 (b), respectively. The X-axis shows the index of the sorted IHPR and the Y-axis shows the IHPR and LIHPR. It can be observed that the IHPR and LIHPR follow a similar trend within a different range of values. Similarly, Figure 4-6 (c) and Figure 4-6 (d) show the decreasing trend of IHPR and LIHPR, respectively. The X-axis shows the index of the sorted IHPR and the Y-axis shows the IHPR and LIHPR. The IHPR and LIHPR also follow a similar decreasing trend. The proposed LIHPR feature reduces the gate count and power consumption by 4.7X and 1.5X, respectively, as compared to IHPR. Therefore, we utilized the LIHPR feature for our negative emotions prediction processor DBE^1.

4.2.3   Feature Normalization

The selected features including PSD, AIHPD, and LIHPR follow a quite different range of values. The ML algorithms can assign higher significance to the features with higher values without feature normalization [74]. Feature normalization is an important ML process in which all the features are transformed into a uniform range. The uniform range of feature values ensures the equal importance of each feature during classifier training [74]. I have used the Z-score feature normalization technique for the feature normalization. The feature vector including sixteen features for the selected eight channels is normalized to a uniform range with zero mean and unit variance using the equation 4.13.

$$NFV = \left[\frac{FV - \mu_{FV}}{\sigma_{FV}} \right] \tag{4.13}$$

FV represents the extracted feature vector, μ_{FV} represents the mean of the extracted features and σ_{FV} represents the standard deviation of the extracted features. The classification results for our negative emotions prediction processor DBE^1 were significantly improved ($\geq 10\%$) after feature normalization. The normalized feature vector (NFV) is then forwarded to the selected classification algorithm for negative emotions prediction.

52
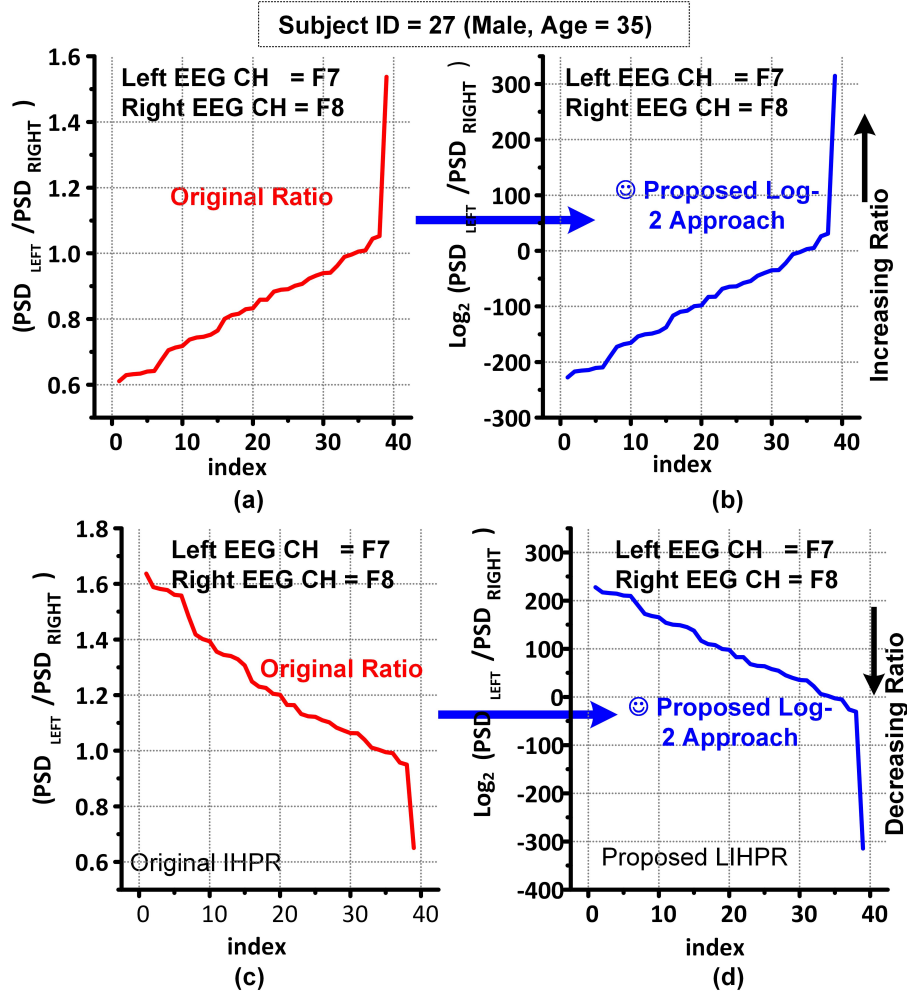
Figure 4-6: (a) Interhemispheric Power Ratio for Increasing Trend (b) Logarithmic Interhemispheric Power Ratio for Increasing Trend (c) Interhemispheric Power Ratio for Decreasing Trend (d) Logarithmic Interhemispheric Power Ratio for Decreasing Trend.

4.2.4 Classifier Selection

A ML classification algorithm is used to predict the label using the selected features. The classifier selection is a composite procedure and depends on a trade-off between classification results and the hardware complexity for on-chip ML processors and hardware applications [38],[39]. The selected features were utilized for negative emotions prediction with different ML classification algorithms. The ML classification algorithms include LR, DT, SVM with linear and RBF kernel, KNN, and NB classifier. Figure 4-7 shows the classification results for negative emotions prediction using DEAP data set with different ML classifiers using

selected features.

The SVM classifier with linear and RBF kernel provided the best classification results. The SVM classifier finds a hyperplane in a k-dimensional space to separate the two classes for binary classification using different kernel functions [75]. The kernel function is used to convert linearly inseparable data to a separable form [76]. The linear SVM classifier consumes less computational energy as compared to the RBF kernel [77]. The on-chip classification parameters storage using the RBF kernel requires $\approx 165X$ memory as compared to the linear SVM classifier using the selected features. Therefore, I selected the linear SVM classifier for the negative emotions prediction DBE processor (DBE^1). The choice of classification algorithm was based on the classification results and the feasibility of on-chip implementation of DBE^1.
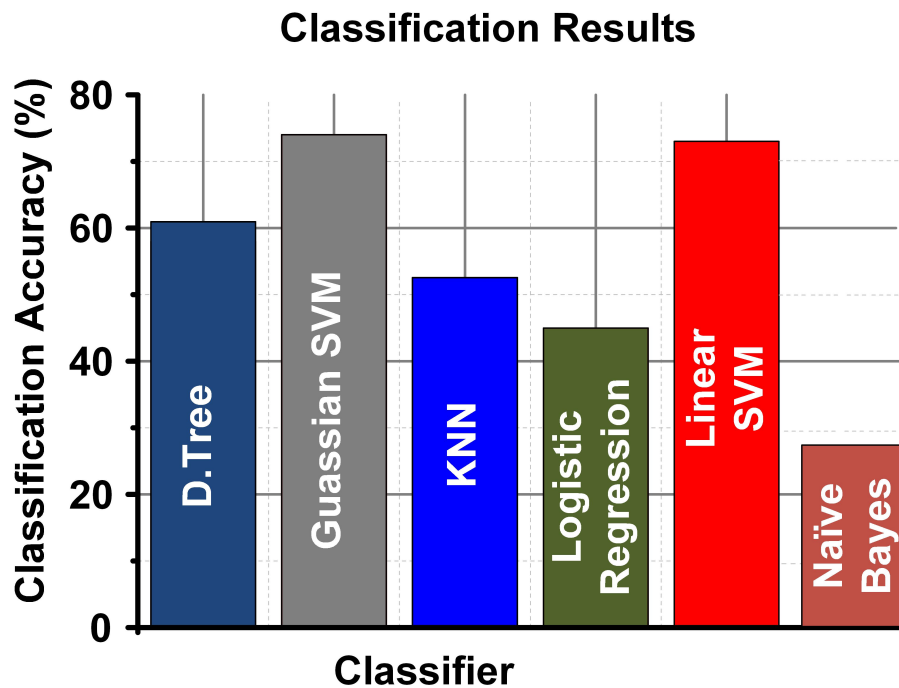
Figure 4-7: Classification Results with Different Machine Learning Classifiers.

4.3 Negative Emotions Prediction Processor Hardware Implementation

The hardware architecture of the proposed negative emotion prediction SoC consists of an AFE, a successive approximation analog-to-digital converter (SAR-ADC), the pre-processing unit (PPU), a feature extraction engine (FEE), and a classification unit (CFU). The AFE utilizing a low-noise capacitive coupled instrumentation amplifier (CCIA) and programmable gain amplifier (PGA) is integrated with DBE^1 processor [27]. The amplified EEG data is digitized by utilizing a 10-bit SAR-ADC, multiplexed among all channels. The digitized EEG signals are then fed to the DBE^1 processor designed for negative emotions prediction.

Figure 4-8 shows the overall architecture of the negative emotions prediction SoC device. The DBE^1 processor is highlighted in a red colored box. The EEG data acquired using AFE is forwarded to the DBE^1 processor for negative emotions prediction. The DBE^1 processor first preprocesses the EEG data to remove noise and other artifacts. The preprocessed data is then forwarded to the FEE. The FEE calculates the PSD, IHPD, and LIHPR features for the selected EEG channels. The features are then forwarded to the feature normalization unit (FNU). The FNU normalizes the extracted features to a uniform range with zero mean and unit variance. The normalized features are then forwarded to the linear SVM CFU for valence and arousal classification, and then negative emotions prediction.

The CFU utilizes the training parameters acquired after offline learning. The parameters are uploaded to the parameter register inside the CFU. The binary valence and arousal classification performed by the CFU is transmitted to a mobile phone using low-power Bluetooth communication module for negative emotions prediction. The low or high valence and arousal labels are used for negative (sad, angry) or positive (relaxed, happy) emotions prediction.

4.3.1 EEG Pre-Processing Unit

The raw EEG data sampled at 4096 samples/second provided by the AFE contains noise and other artifacts (eye blinking, chewing, etc.). The EEG data is then down-sampled to 128 Hz using a programmable down-sampling unit (DS) unit. To minimize these artifacts and noise, a mean average referenced (MAR) unit and a band-pass filter from 2-60 Hz are utilized.
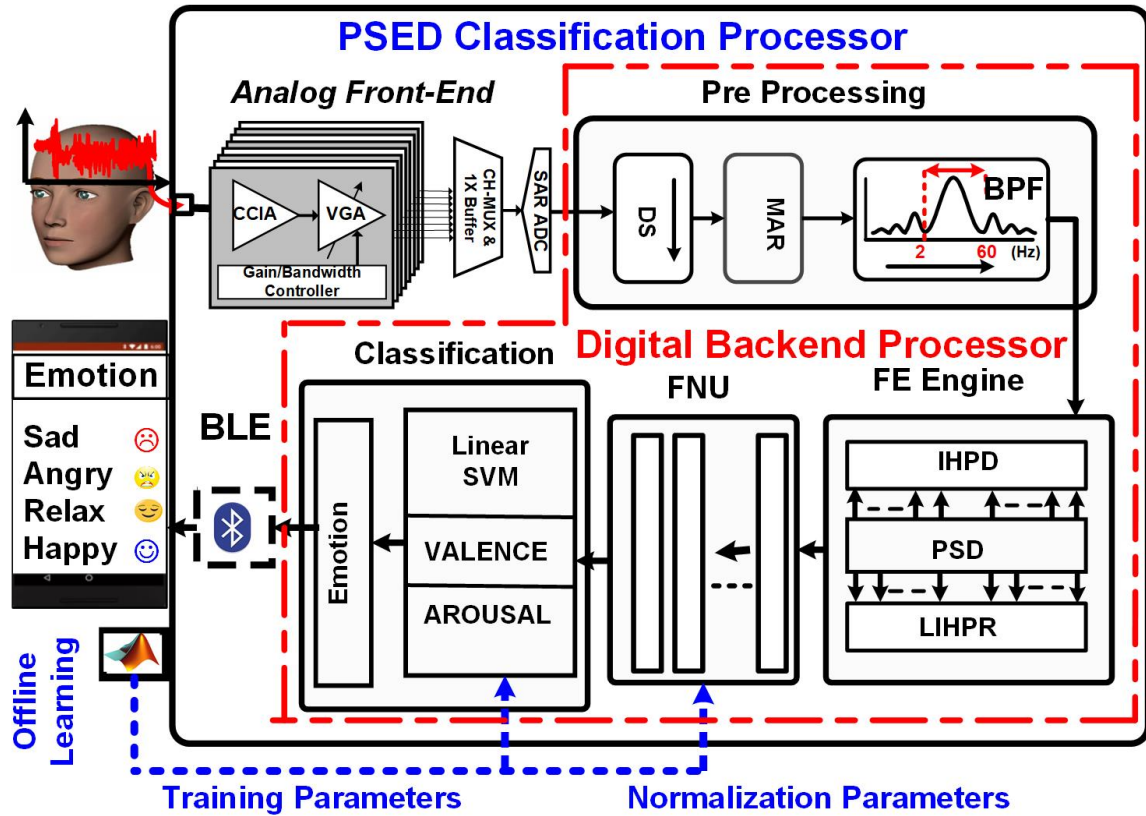
Figure 4-8: Negative Emotions Prediction Processor Architecture (DBE^1).

4.3.2 Feature Extraction Engine

The FEE is used to calculate the selected features including PSD, AIHPD, and LIHPR from preprocessed EEG signals. Figure 4-9 shows the hardware architecture of the FEE. The calculated features $(FV_1 - FV_{16})$ are forwarded to the feature normalization unit for Z-score normalization. The normalized features $(NFV_1 - NFV_{16})$ are then forwarded by the feature extraction unit for negative emotions prediction to the CFU.

The EEG data of the selected eight channels is forwarded to the FEE. A 3-bit channels selection input (ch-sel) is used to indicate the selected channel. A 3-bit comparator followed by a 3-bit flip flop indicates the change in the channel to reset the feature calculation. The incoming EEG data of the eight channels are sequentially forwarded to the FEE. This utilized 8X times lower area and input pins than a parallel implementation. A bandpass filter is used to calculate the PSD feature in the beta band (12-30 Hz). The bandpass filter is implemented using a finite impulse response filter of the 50th order utilizing the least

Figure 4-9: Feature Extraction Engine Hardware Architecture.

square method to minimize the number of filter coefficients. The passband and stopband frequencies of the filter are 0db and -43db, respectively, to maximize the signal-to-noise ratio (SNR). PSD is calculated by accumulating the absolute values of the band-passed EEG signal using a 32-bit integer accumulator and absolute unit (ACCM-ABS). The output of the ACCM-ABS is quantized to 16 bits and stored in a memory block (MEM-PSD) of 8 x 16 bits as P1-P8. The ch-sel provides the address input for the MEM-PSD.

The PSD for the four right and left hemisphere pairs $(P_1 - P_2, P_3 - P_4, P_5 - P_6, P_7 - P_8)$ are forwarded for the AIHPD, and LIHPR calculation using two 4-to-1 multiplexers and ch-sel. AIHPD and LIHPR are calculated by absolute subtraction (ABS-SUB) and logarithmic division units (LDU), respectively as shown in Figure 4-9.

AIHPD and LIHPR Implementation

AIHPD is calculated using a 16-bit subtractor and substituting the sign bit with zero. AIHPD is stored in a memory block (MEM-AIHPD) of 4 x 16 bits as D1-D4 using ch-sel as address input. The hardware implementation of the LIHPR feature requires a logarithmic unit to calculate the logarithm function. The logarithm function is usually implemented by a LUT, piecewise approximation, or CORDIC algorithm [78].

Figure 4-10 shows the architecture of the LDU for LIHPR calculation. The PSD values of the selected channels follow a different range. The conventional implementation of the logarithmic unit using LUT for LIHPR calculation requires a huge ($> 1MB$) amount of memory. Therefore, the PSD values of the selected channels are down-scaled and rounded off to a uniform range (0-1023) using a feature downscaling and rounding unit. Equation 4.14 is used to downscale and round the features to a uniform range.

$$P_{DS} = \left[(P - P_{MIN}) * \frac{R_{MAX} - R_{MIN}}{P_{MAX} - P_{MIN}} + R_{MIN} \right] \qquad (4.14)$$

P represents the PSD value, and P_{DS} represents the downs scaled PSD value. R_{MAX} and R_{MIN} are the maximum and minimum values of the targeted range of values, which are 0 and 1023, respectively, in this case. P_{MAX} and P_{MIN} represent the maximum and minimum PSD values for each channel. Since R_{MAX} and R_{MIN} are constant numbers, equation 4.14 was simplified as equation 4.15 and then 4.16 for hardware implementation. $R_{MXMNDIFF}$ and $P_{MXMNDIFF}$ are the differences of R_{MAX}, R_{MIN}, and P_{MAX}, P_{MIN}, respectively in equation 4.15. $RP_{MXMNDIFF}$ is the ratio of $R_{MXMNDIFF}$ and $P_{MXMNDIFF}$ in equation 4.16. Equation 4.16 is expanded as 4.17, which is further simplified as equation 4.18.

$$P_{DS} = \left[(P - P_{MIN}) * \frac{R_{MXMNDIFF}}{P_{MXMNDIFF}} + R_{MIN} \right] \qquad (4.15)$$

$$P_{DS} = \left[(P - P_{MIN}) * RP_{MXMNDIFF} + R_{MIN} \right] \qquad (4.16)$$

58

$$P_{DS} = \left[P * RP_{MXMNDIFF} - (P_{MIN} * RP_{MXMNDIFF} + R_{MIN}) \right] \quad (4.17)$$

$$m \qquad\qquad P_{DS} = \left[(P * SP_x) - SP_y \right] \qquad\qquad (4.18)$$

$$P_{DSR} = \left[round(P_{DS}) \right] \qquad\qquad (4.19)$$
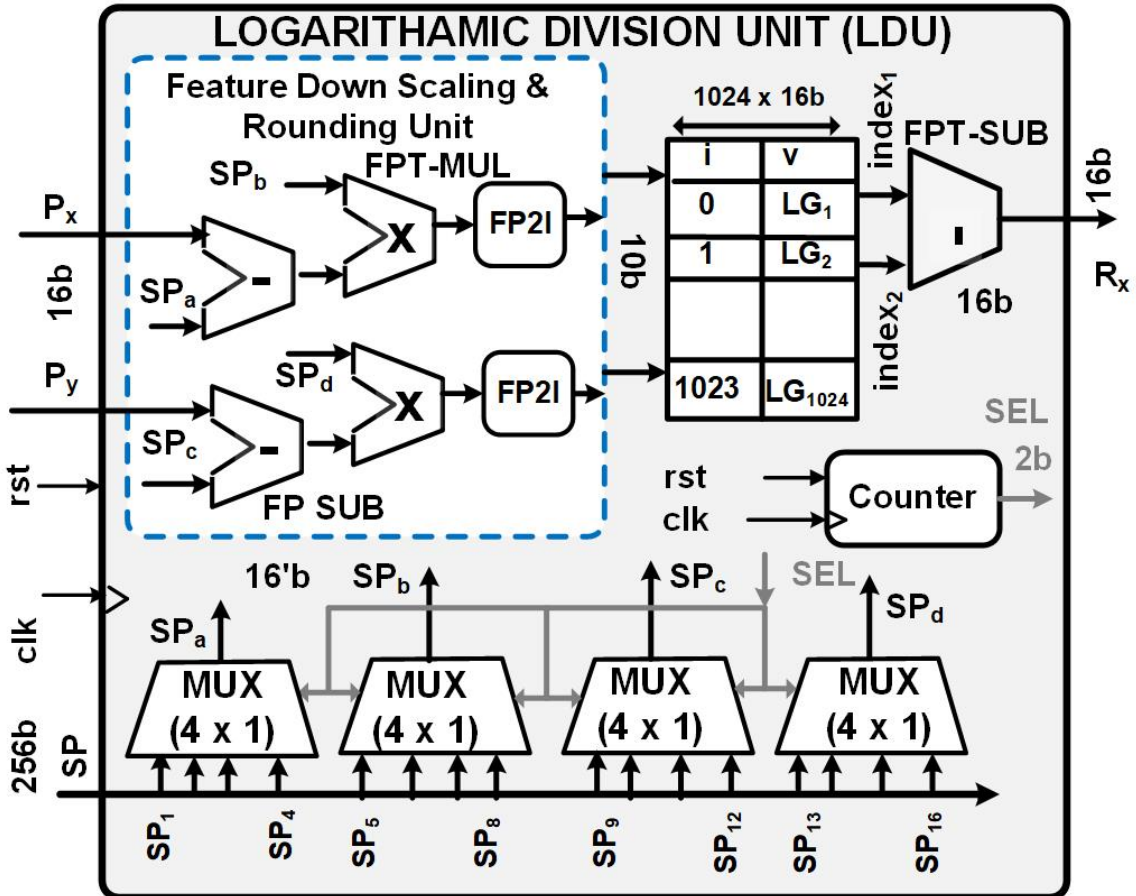
Figure 4-10: Logarithmic Division Unit for Logarithmic Interhemispheric Power Ratio Calculation.

SP_a and SP_b are the scaling parameters for each EEG channel. These scaling parameters

59

are uploaded to the parameters register along with classification parameters. Sixteen scaling parameters $(SP_1 - SP_{16})$ are uploaded for the selected subset of eight channels.

The LIHPR of PSD in comparison to LIHPR of down-scaled and rounded-off PSD values (PSD_{DSR}) is shown in Figure 4-11. The $LIHPR(PSD_{DSR})$ has a much narrower and smaller range as compared to the $LIHPR(PSD)$ and appears as a constant value as shown in the top part of Figure 4-11. However, if the value is zoomed in, it can be observed that it follows a similar trend in a narrower range (0-80). The graph of PSD_{DSR} in the smaller range can be seen in the bottom part of Figure 4-11. Therefore, we have utilized the $LIHPR(PSD_{DSR})$ instead of $LIHPR(PSD)$ for our negative emotions prediction DBE processor DBE^1.

The range of bits for PSD_{DSR} has a direct impact on the hardware resources and the classification results. If the PSD values are down-scaled and rounded off to a wider range, the $LIHPR(PSD_{DSR})$ provides a better approximation to the actual LIHPR (PSD). However, it would also utilize higher hardware resources. I, therefore, analyzed the classification results along with a variable number of bits for the LUT storage to calculate LIHPR. Figure 4-12 shows the effect of quantization of the number of bits of LUT for LIHPR implementation. The valence and arousal prediction results against the variable size of the LUT to calculate $LIHPR(PSD_{DSR})$ is shown in Figure 4-12 (a). It was observed that the classification results were not affected if the number of bits for the LUT index is reduced from 16 bits (64k elements) to 10 bits (1k elements). However, if the number of bits for the LUT index is further reduced, the classification results are degraded ($> 5\%$). Therefore, a LUT with 10 bits index (1024 elements) is used for the $LIHPR(PSD_{DSR})$ calculation. This quantization of the number of bits for the LUT index reduces the memory requirement for LUT from 1MB to 16KB as depicted in Figure 4-12 (b). The proposed LDU for LIHPR calculation replaces the division operation with 34% lesser area compared to piece-wise linear and CORDIC algorithms for the division. But, the proposed implementation cannot be used for the general division and is customized for this ML application.

Feature Normalization Unit

The feature normalization unit in Figure 4-9 calculates the normalized features as defined in equation 4.13. The μ_{FV} and $1/\sigma_{FV}$ of the selected features are acquired from the parameter registers as normalization parameters N_1-N_{16} and N_{17}-N_{32}, respectively. Two
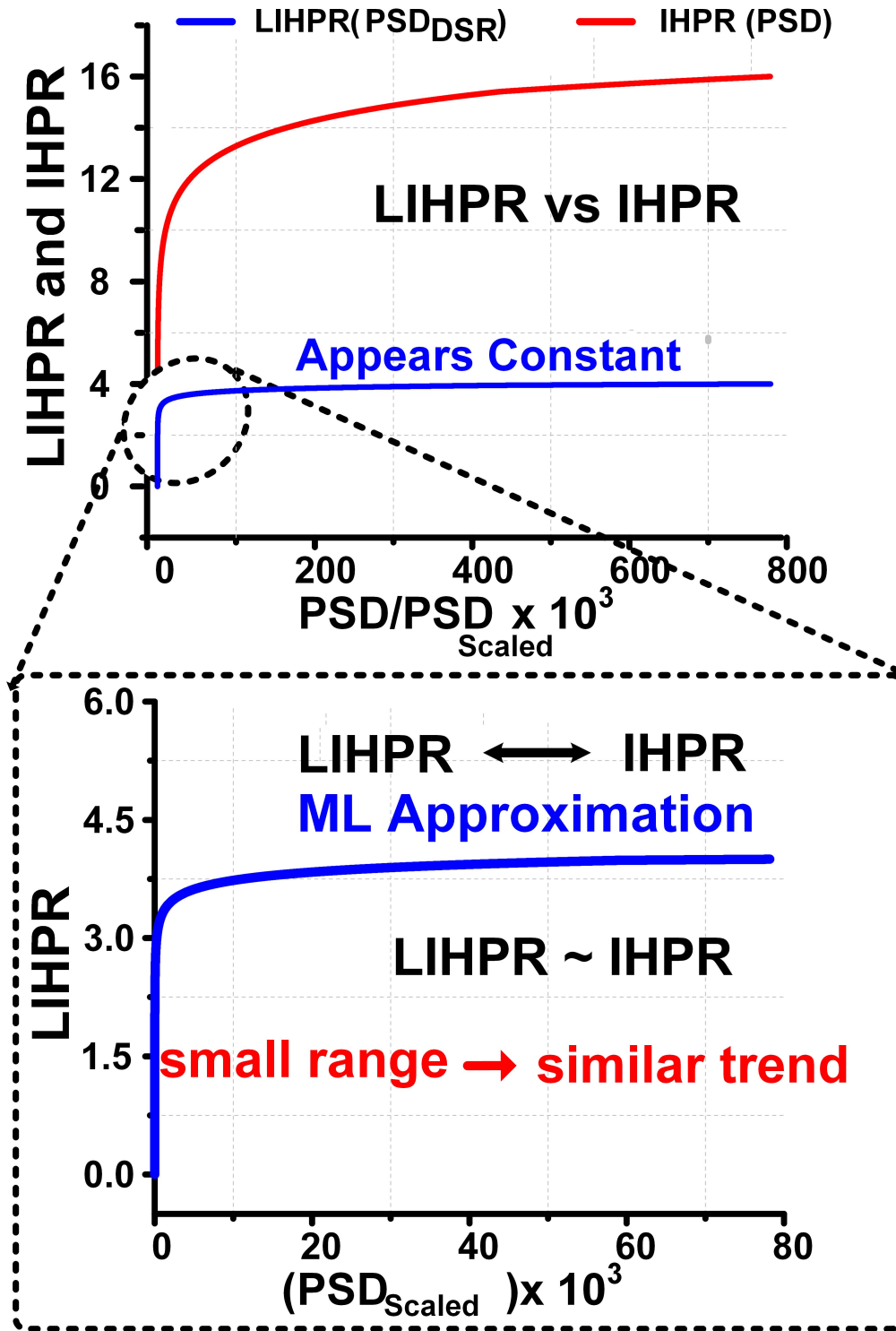
Figure 4-11: Logarithmic Interhemispheric Power Ratio (Power Spectral Density) & Proposed Logarithmic Interhemispheric Power Ratio (Down Scaled and Rounded Off).
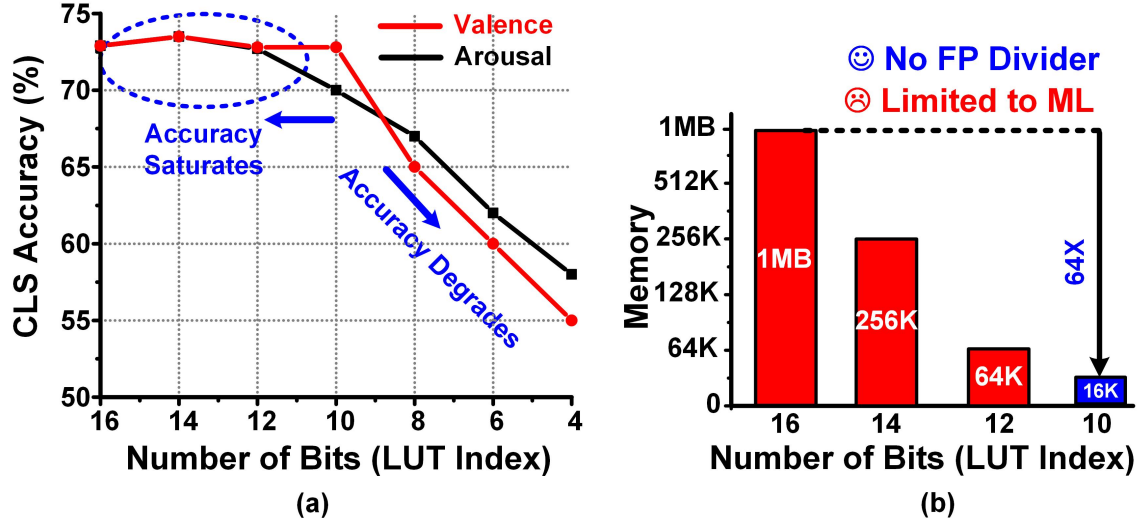
Figure 4-12: (a) Number of Bits for Look Up Table Index & Classification Results (b) Number of Bits for Look Up Table Index & Memory Requirement.

16-to-1 multiplexers are used to select the normalization parameters N_a and N_b for each feature. A 4-bit counter is used for the selection input of these multiplexers.

The normalized feature vector (NFV_{1-16}) is stored in a memory block (MEM-FV) of 16 x 16 bit FLP numbers and then forwarded to the CFU for negative emotions prediction. The CFU utilizes the normalized features along with patient-specific and pre-uploaded classification parameters for negative emotions prediction.

4.3.3   Classification Parameters Upload

Patient-specific or subject-dependent classification parameters are uploaded to a parameters register for the negative emotions prediction. The parameter uploading is performed using write-enable (WE), strobe (STB), write data in (WDIN), write data out (WDOUT), and write done (WDONE) inputs and outputs in negative emotions prediction SoC. The WEN and STB initiate the parameters uploading serially through WDIN. The WDONE indicates the completion of parameters uploading.

Figure 4-13 shows the parameter uploading process which starts when the STB and WEN inputs are asserted after reset. Figure 4-13 (a) shows the state diagram and the timing diagram for the parameters upload. The default parameters are uploaded to the parameters register if the user does not want to upload new parameters. The parameters
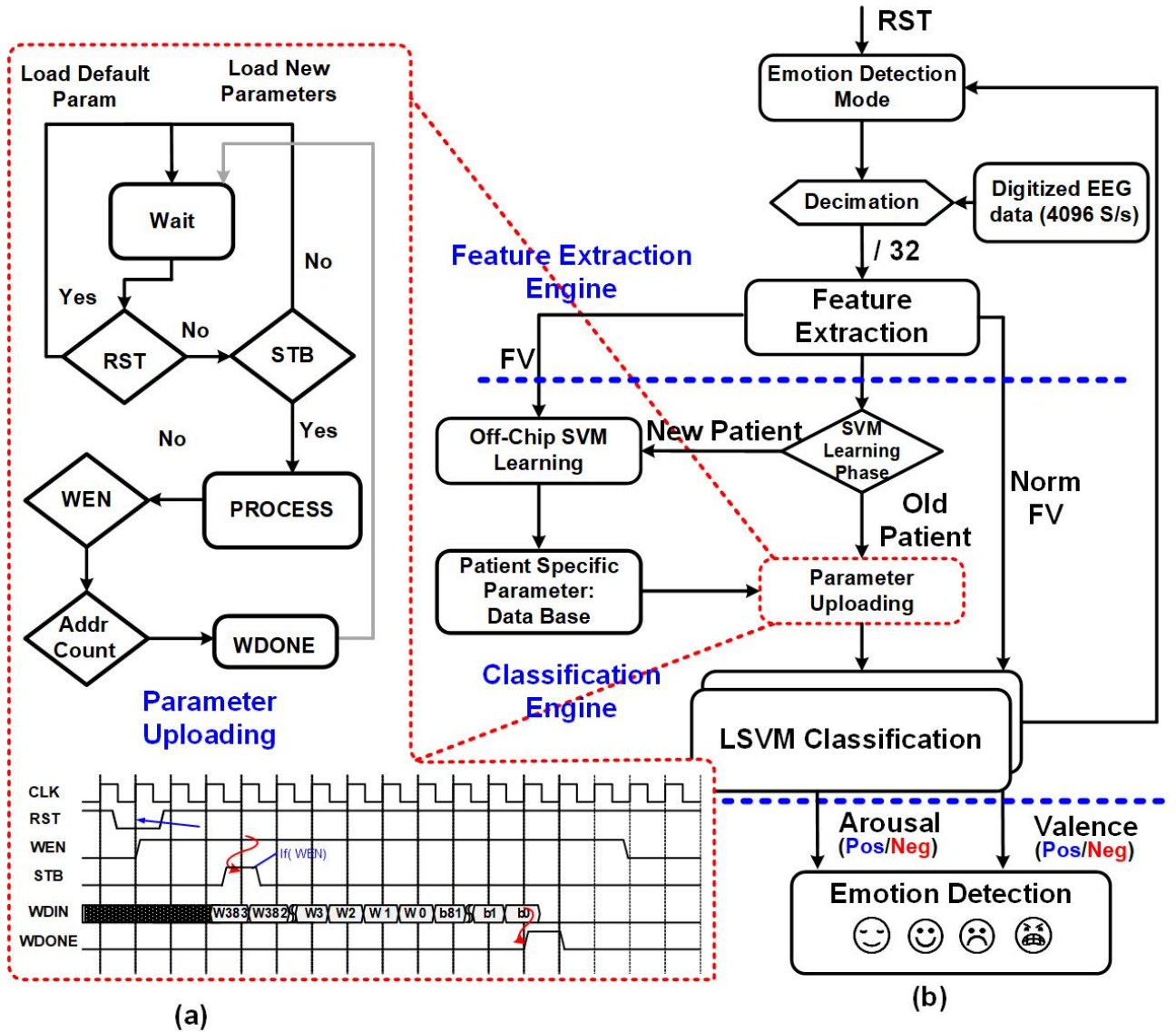
Figure 4-13: (a) State & Timing Diagram for Parameter Uploading (b) State Diagram for Emotions Prediction Processor.

are forwarded serially through WDIN input after asserting STB and WEN inputs. The WDONE output indicates the completion of the parameter uploading process.

Figure 4-13 (b) shows the complete EEG-emotion classification process by a state diagram. The digitized and decimated EEG data is forwarded to the FEE to calculate the feature vector. The classification parameters are acquired from the parameter register and uploaded serially as explained earlier. The feature vector and the classification parameters are forwarded to the linear SVM CFU for negative emotions prediction using valence and

arousal classification. The valence and arousal classes determine the negative (sad, angry) or positive (happy, relaxed) emotions.

4.3.4   Classification Unit Implementation

The CFU calculates the labels of valence and arousal as positive (one) or negative (zero) using a linear SVM classifier. The NFV_{1-16} and pre-loaded classification parameters (VAL-P and ARR-P) for valence and arousal, respectively, are forwarded to the CFU by FEE and parameter register, respectively. A 40 kHz clock (iCLK) is used to select the VAL-P or ARR-P using a 2 x 1 multiplexer. An LSVM classifier is utilized to ensure area-and-power-optimized hardware implementation.

Linear SVM performs a binary classification using weights (W), the scaled feature vector (H), and bias (B), defined in equations 4.20-4.23. H_{1-16} is obtained by adding the scale shift (SS_{1-16}) to the normalized feature vector (NFV_{1-16}) and then multiplying the intermediate vector named K_{1-16} with scale factor (SF_{1-16}) as in equations 4.20 and 4.21. The SS_{1-16}, SF_{1-16}, W_{1-16} and, B are extracted from the 784-bit valence and arousal classification parameters (VAL-P or ARR-P) as depicted in Figure 4-14.

$$K_{1-16} = \left[\ NFV_{1-16} + SS_{1-16} \ \right] \tag{4.20}$$

$$H_{1-16} = \left[\ K_{1-16} * SF_{1-16} \ \right] \tag{4.21}$$

$$Y_{1-16} = \left[\ \sum_{n=0}^{16} W_n.H_n + B \ \right] \tag{4.22}$$

$$Valence/Arousal = \begin{cases} 1, & \text{if } Y < 0 \\ 0, & \text{if } Y \geq 0 \end{cases} \tag{4.23}$$
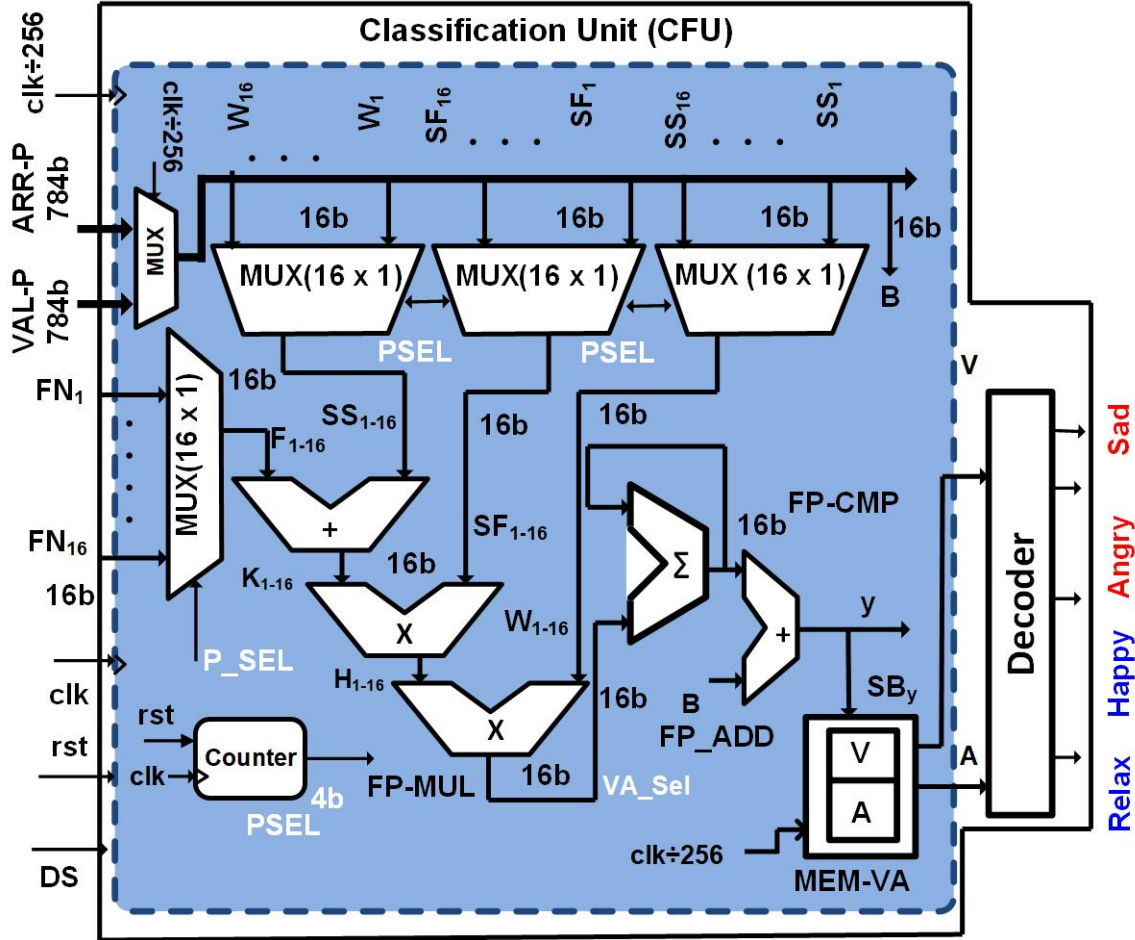
Figure 4-14: Classification Unit Hardware Implementation.

Figure 4-14 shows the architecture of the hardware unit for linear SVM classifier implementation. Two 16-bit FLP multipliers (FPT-MUL), one 16-bit FLP adder (FPT-ADD) and one 16-bit FLP accumulator (FPT-ACCM) are used to implement the equations 4.20-4.22. Three 16 x 1 multiplexers are used to select the SS_x, SF_x, W_x, and H_x from SS_{1-16}, SF_{1-16}, W_{1-16}, and H_{1-16}, respectively. The selection input (P-SEL) of the multiplexer is derived by a 4-bit counter.

The valence and arousal label is calculated using a sign bit of Y (SB_Y) in equation 4.23. The SB_Y is stored in a memory block (MEM-VA) of 2 x 1 bit for valence and arousal classification. Finally, negative or positive emotion is determined using a 2-to-4 decoder based on the valence and arousal classification.

65

4.4 Classification Performance & Measurement Results

The classification performance of the proposed negative emotions prediction processor DBE^1 is evaluated using classification accuracy, sensitivity, and specificity. The classification accuracy represents the ratio of correct predictions. It is calculated from the number of true positive (TP), true negative (TN), false positive (FP), and, false negative (FN) predictions performed. The TP represents the classes that are positive in actual and are truly predicted as positive by the classifier. The TN represents the classes that are negative in actual and are predicted negative (truly) by the classifier. The FP represents the classes that are actually negative and are predicted positive (falsely) by the classifier. The FN represents the classes that are actually positive and are predicted negative (falsely) by the classifier. The formulas for the calculation of classification accuracy, specificity, and sensitivity are shown in equations 4.24, 4.25, and 4.26, respectively.

$$Classification\ Accuracy = \left[\frac{TP + TN}{TP + TN + FP + FN} \right] \tag{4.24}$$

$$Sensitivity = \left[\frac{TP}{TP + FN} \right] \tag{4.25}$$

$$Specificity = \left[\frac{TN}{TN + FP} \right] \tag{4.26}$$

The summary of the overall classification results is shown in Table 4.2. The classification accuracies of 72.96% and 73.14% were obtained for the DEAP data set for valence and arousal classification, respectively. The classification accuracy of 70.71% was obtained for the valence classification using the SEED data set. The best-case patient-specific classification accuracy reported for DEAP and SEED data sets are $\approx 90\%$ and 100%, respectively. The sensitivity and specificity of 73.0% and 73.1%, respectively, are achieved using the DEAP data set, whereas, the sensitivity and specificity of 70% and 71.4%, respectively, are achieved using the SEED data set.

Table 4.2: Classification Results for DBE^1 Negative Emotions Prediction Processor.

Data Set	Valence	Arousal
DEAP	72.96%	73.14%
SEED	70.71%	——

Figure 4-15 shows the patient-specific or subject-dependent classification results for DEAP data set. The classification results of all subjects in the DEAP data set for valence and arousal classification are shown in Figure 4-15 (a) and 4-15 (b), respectively. The subject-wise patient-specific classification results of this negative emotions prediction DBE processor DBE^1 using the SEED data set are shown in Figure 4-16.

The classification results are evaluated using the leave-one-out cross-validation method. Cross-validation is the process of validating the trained ML classification model against unknown EEG data to avoid over-fitting. K-fold and leave-one-out validation methods are most widely used for emotion classification processors [59]. The leave-one-out cross-validation scheme trains the classifier on all emotion labels except one label and repeats this process for all labels of a subject. I chose this validation scheme because of its lower dependence on the data partitioned during the train and test split and a variable number of positive and negative classes for each subject.

4.4.1 Measurement Results

The measurement results represent the testing of the proposed negative emotions prediction processor. Figure 4-17 demonstrates the measurement result of two scenarios of negative and positive emotions. The negative and positive emotions correspond to negative valence with positive arousal and positive valence with positive arousal corresponding to angry and happy emotions, respectively. It shows the 15 seconds of selected eight EEG channels for the trials (subject # 8) from the DEAP data set.

The EEG signals of the selected channels are forwarded to the negative emotions prediction processor. A feature vector compromising PSD, AIHPD, and LIHPR is calculated by the FEE. The calculated features are then Z-score normalized by the feature normaliza-
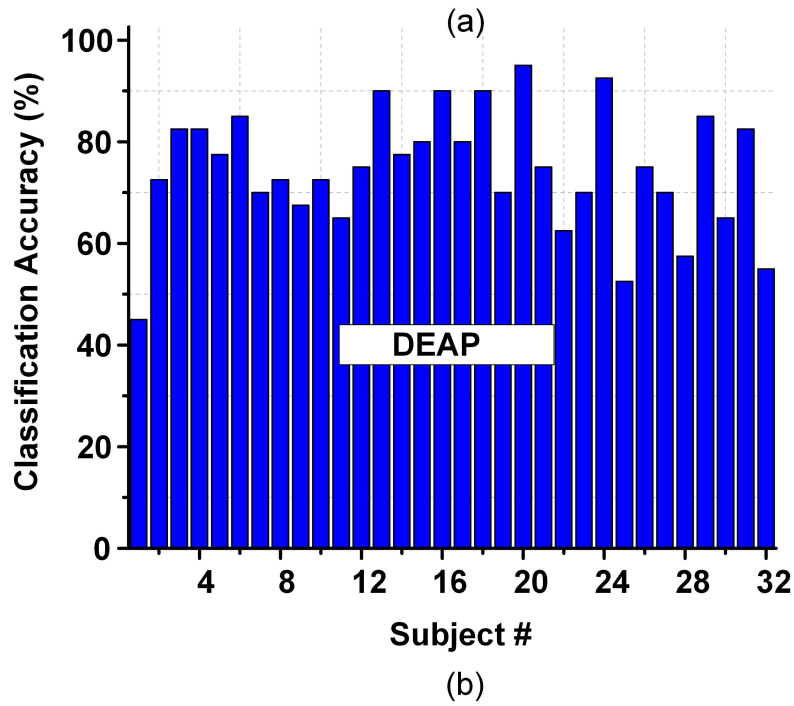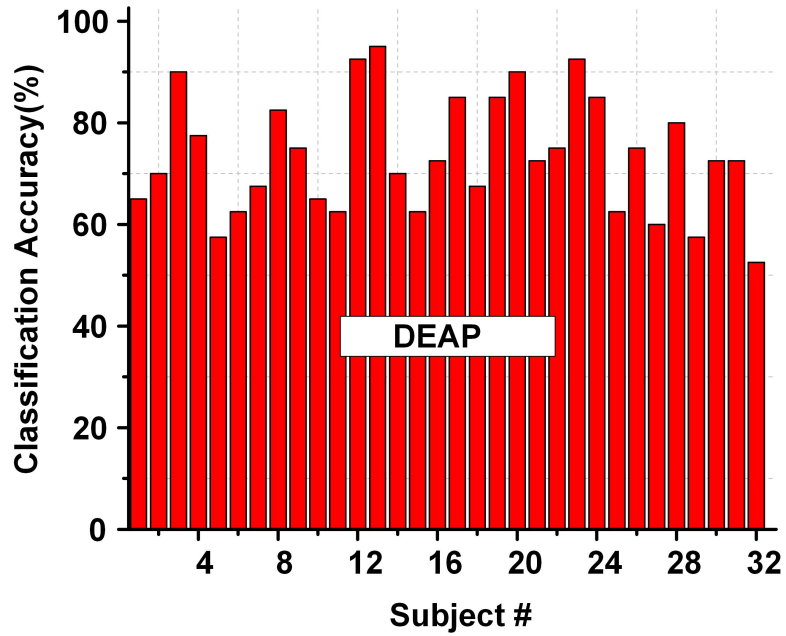
Figure 4-15: Subject-Wise Classification Results using DEAP for (a) Valence Classification (b) Arousal Classification.

tion unit. The normalized feature vector is forwarded to the linear SVM CFU for positive or negative emotions prediction using valence and arousal labels. The classification of low
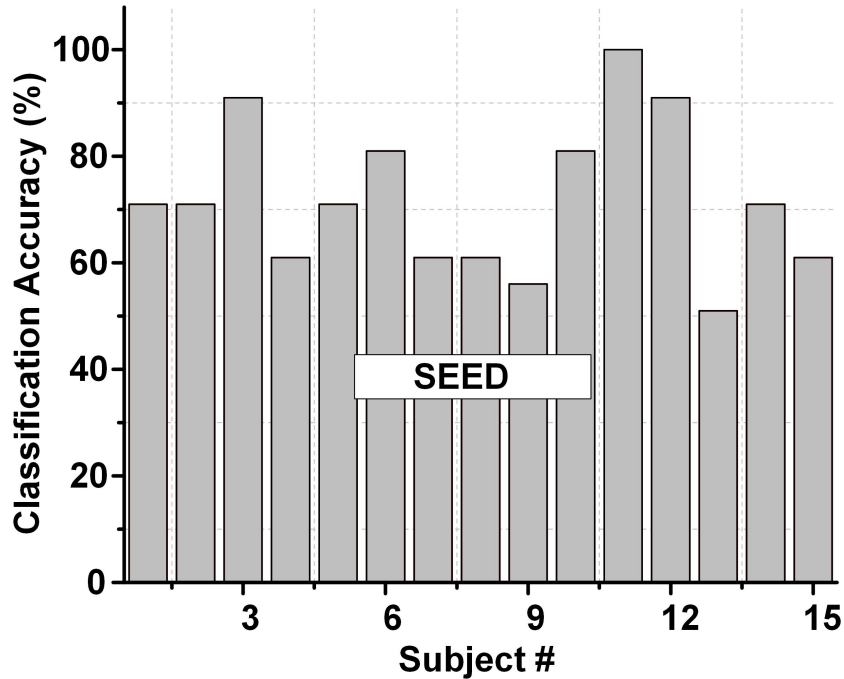
Figure 4-16: Subject-Wise Classification Results using SEED Data Set.

valence and high arousal (sad emotion) matches the original labels of low (1.21) valence and high (7.94) for valence and arousal, respectively as shown in Figure 4-17 (a). Similarly, another EEG segment corresponding to a happy emotion (high valence, high arousal) is shown in Figure 4-17 (b). The classification of high valence and high arousal matches the original labels of high (8.15) valence and high (6.96) arousal.

4.4.2 Chip Performance Summary

The chip performance summary summarizes the performance of the SoC implementation of this proposed negative emotions prediction processor. DBE^1 negative emotions prediction DBE processor is fabricated using TSMC $0.18\mu m$ 1-Poly-6-Metal (1P6M) CMOS technology. The chip micrograph and the performance summary of DBE^1 are shown in Figure 4-18.

The DBE^1 was implemented with an active area utilization of $5.4mm^2$ including the area utilization of $3.6mm^2$, $1.4mm^2$, and $0.4mm^2$ for the FEE, CFU, and the support vector cache (SVC), respectively. The support vector cache represents the parameters register for classification parameters. It stores the classification parameters. The classification
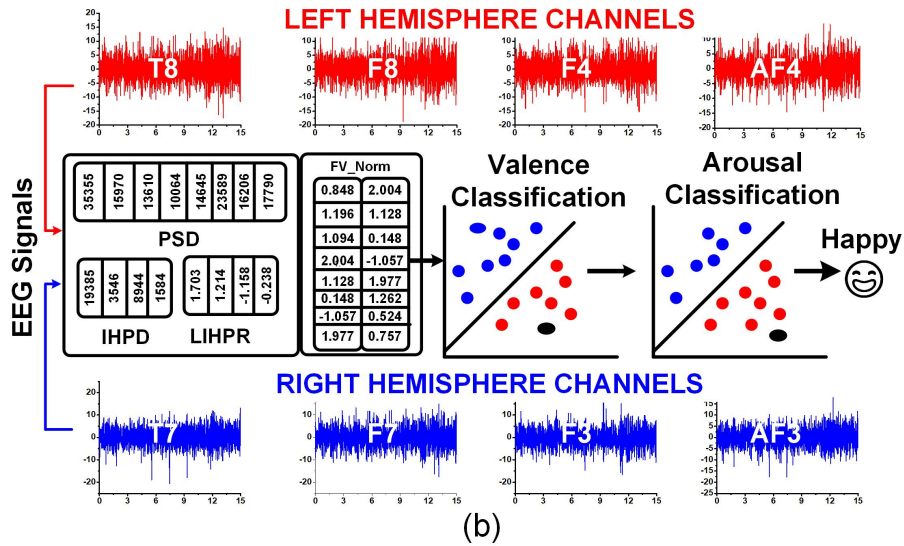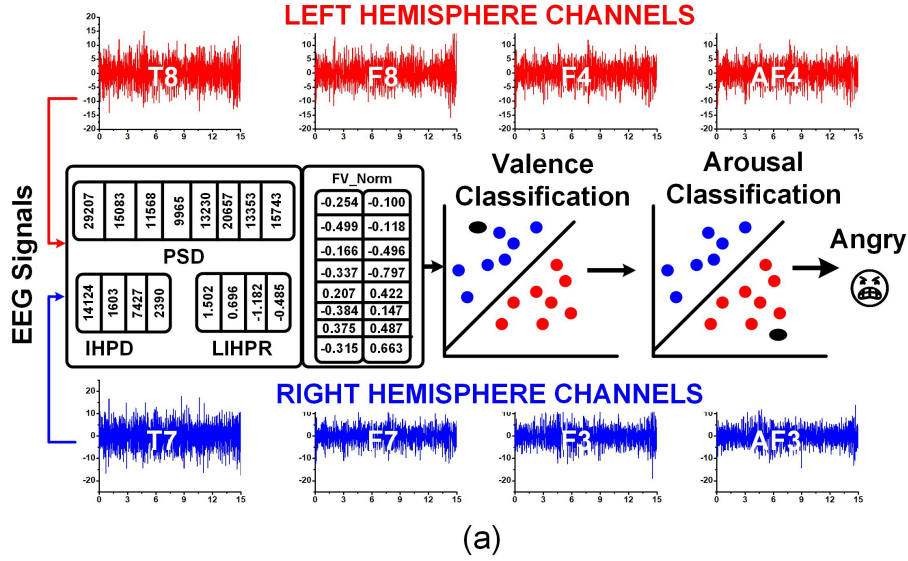
Figure 4-17: DBE^1 Negative Emotions Prediction Processor Measurement Results for (a) Angry (b) Happy Emotion on a Subject in DEAP.

parameters are obtained during classifier training (offline) and are uploaded to the system as explained earlier. The supply voltage of 1.0 V was used for this SoC implementation. The proposed SoC performs the negative emotions prediction with the power and energy utilization of 2.04mW and $16\mu J/prediction$, respectively using eight EEG channels.
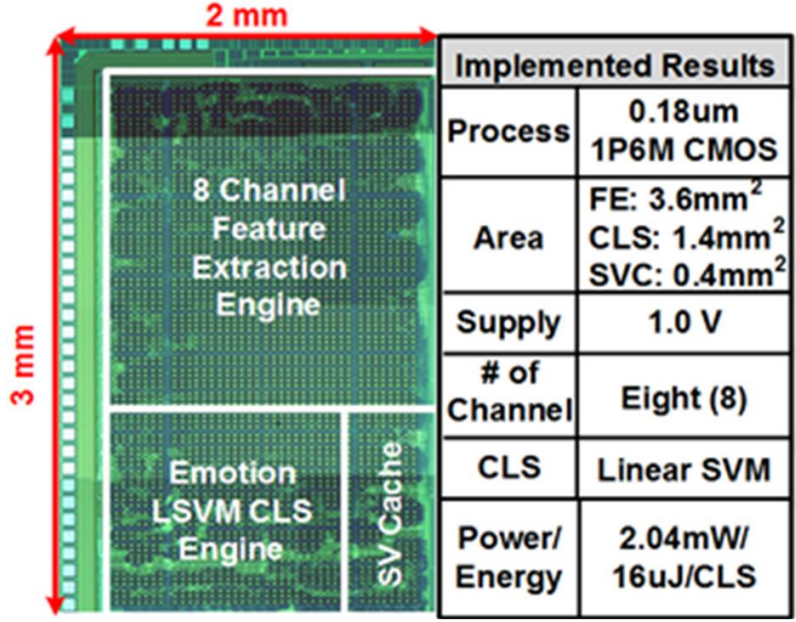
Figure 4-18: Chip Micrograph & Performance Summary of System.

Classification Energy and Battery Time Calculation

The total power consumption of the DBE^1 synthesized on a 1 MHz clock frequency is 2.04 mW, which down-scales to $\approx 2088\mu W$ with a 1 kHz clock frequency. The power of a system is defined as the energy per unit of time. Therefore, the energy utilization of the DBE^1 per prediction can be calculated by multiplying the power by the number of time units consumed per prediction, which is approximately $16\mu J/prediction$. The total estimated backup time of a battery can be calculated as equation 4.27. The estimated battery time in hours for DBE^1 processor using a coin cell battery of size, voltage, and area of 20 mm, 3 V, and 250 mAH, respectively is calculated in equation 4.28, which is ≈ 367 Hrs. Considering the aging factor and other losses of 50%, the estimated time would be ≈ 183 Hrs or 7 days.

$$Estimated\ Backup\ Time(Hrs) = \left[\frac{Battery\ Capacity\ X\ Input\ Voltage}{Total\ Load}\right] \quad (4.27)$$

$$Estimated\ Backup\ Time(Hrs) = \left[\frac{250mAH\ Capacity\ X\ 3V}{2088\mu W}\right] \quad (4.28)$$

Table 4.3: Comparison With the State-of-The-Art Works.

	S. Koelstra	H. Ullah	W.Fang	W.Zheng	H. Soleymani	DBE^1
	[24]	[79]	[26]	[68]	[49]	[38]
Hardware	No	No	Yes (28nm)	No	No	**Yes**$(0.18\mu m)$
Area	—	—	$3.35mm^2$	—	—	**5.4** mm^2
Power/Energy	—	—	76.61mW	—	—	$16\mu J$
Data Set(s)	DEAP	DEAP	DEAP	DEAP,SEED	MAHNOOB	**DEAP, SEED**
Ch. Count	32	Variable	6	32	32	**8**
Classifier	NB	SVM	CNN	GELM	LSTM	**SVM**
Accuracy (%)	59.8	73.8	80.1	80.4	82.5	**74**
Multiple BM-DS	No	No	No	Yes	No	**Yes**

4.4.3 Performance Comparison

The proposed DBE^1 DBE processor was the first fully integrated on-chip processor for negative emotions prediction. The performance comparison of the DBE^1 negative emotions prediction processor with the previous state-of-the-art works is summarized in Table 4.3.

S.Koelstra [24] performed the emotions classification using 32 EEG channels. They achieved the average classification results of 59.8% using a naive Bayes classification algorithm and DEAP data set. They did not provide any hardware implementation of the system and the system was software-based. The system was validated on only one data set (DEAP), and the algorithm was not validated on multiple BM-DS.

H. Ullah [79] reported the average classification results of 73.8% for the emotions prediction using a linear SVM classifier and DEAP data set. They have used an automatic channel selection mechanism, with a variable number of EEG channels, which is not suitable for hardware implementation. They also lack multiple BM-DS validations.

W. Fang [26] is the only other hardware-based implementation for an emotions prediction system targeting an SoC. The system was implemented using a 28nm CMOS technology process. They achieved the average classification results of 80.1% for emotions prediction

using a convolutional neural network classifier and DEAP data set. They utilized only 6 EEG channels for emotion prediction. The area and power utilization by this system were $3.35mm^2$ and 76.61mW, respectively. However, they did not provide a fully integrated SoC or DBE processor. The features are calculated offline and then forwarded to the SoC for emotion prediction. They also lack multiple BM-DS validations.

W.Zheng [68] proposed an emotions classification algorithm using DEAP and SEED data sets. They achieved the average classification accuracy of 80.4% using a GELM classifier [80]. Although, they have validated their system on multiple BM-DS. But, they also lack a hardware implementation and have utilized a large number of EEG channels (DEAP=32, SEED=64).

H. Soleyemani [49] provided the average classification result of 82.5% on a self-recorded private data set (MAHNOOB) using LSTM classification. They have not utilized any BM-DS to validate their emotions classification algorithm. They have also utilized a large number (32) of EEG channels, and lack hardware implementation.

In this, first, fully integrated SoC implementation for a DBE processor (DBE^1) for negative emotions prediction, a very good average classification accuracy of $\approx 74\%$ was achieved for multiple BM-DS including DEAP and SEED data sets. The DBE^1 utilizes only eight EEG channels with minimum power (2.04 mW), classification energy ($16\mu J/prediction$), and area ($3.35mm^2$) with a fully integrated SoC solution.

4.5   Summary

In this chapter, I have explained the design of DBE^1 negative emotions prediction processor. The DBE^1 negative emotions prediction processor has the following novel points:-

- 1st fully integrated SoC-based 8-channel EEG-based emotion classification processor.

- Utilizes hardware-efficient linear SVM classifier.

- Can assist in the learning and cognitive development of ASD patients.

- Continuously detects human emotions using valence and arousal scales.

- Validated using well-known BM-DS including DEAP and SEED.

- Utilizes area-and-energy-efficient features including PSD, AIHPD, and LIHPR.

- Proposed approximated implementations of the FV with 4.7X reduced area.

- The LUT-based log-2 divider utilized for LIHPR is 34% more area-efficient.

- Provides very good classification results on DEAP (73.14%) and SEED (75.8%).

- Fabricated using TSMC 180nm CMOS process

- Consumes an overall classification energy of $16\mu J$ at 1 kHz with a system latency of < 0.8 min for every emotion prediction.

4.6 Conclusion

The proposed DBE^1 negative emotions prediction processor provided very good (74%), but not excellent ($< 80\%$) classification results for negative emotions prediction. The system was fully validated on DEAP data set and on randomly selected sessions in SEED data set. The classification energy ($16\mu J/prediction$) and the number of EEG channels (8) for the negative emotions prediction was also high. In order to overcome these shortcomings, a second-generation processor DBE^2 was proposed and developed. The DBE^2 processor utilizes deep neural networks for emotion prediction. The detailed algorithm and hardware architecture of DBE^2 are explained in the next chapter (Chapter 5).

Chapter 5

DBE^2: 2nd Generation DBE Processor for Negative Emotion Outburst Prediction

The design of the 2nd generation DBE processor (DBE^2) is presented in this chapter. The challenges related to the processor design for the suitable channels, features with classifier identification, and hardware implementation are discussed. The classification algorithm, hardware implementation, and optimization of the identified features with the classification algorithm to minimize the area, and classification energy of the proposed DBE^2 are explained in this chapter. Section 5.1 explains the process of emotion prediction using EEG signals. The choice of data sets, labels for emotion prediction, and the challenges in emotion prediction using EEG signals is explained in this section. Section 5.2 explains the ML algorithm used for negative emotions prediction in DBE^2. The selection of a suitable subset of EEG channels, features, and classification algorithm selection are explained in this section. Section 5.3 explains the hardware architecture for the SoC implementation of the selected ML algorithm in DBE^2. The overall architecture of DBE^2, hardware implementation of EEG pre-processing unit, feature extraction engine including feature normalization unit, classification unit, and the parameters register for classification parameters upload is explained in this section. The classification performance of DBE^2 including classification results, chip performance summary, and performance comparison with the other state-of-

the-art works are discussed in section 5.4. Section 5.5 summarizes the overall achievements and shortcomings of DBE^2.

5.1 Introduction

A primary challenge for ASD patients is emotional dysregulation and NEOB as discussed earlier in chapter 1 and chapter 2. The early prediction of these NEOB's can therefore be used to control their effects by regulating emotion. The DBE^2 NEOB prediction DBE processor was therefore designed and developed. The DBE^2 addresses and provides a solution to the shortcomings of previous solutions including DBE^1. The shortcomings of previous solutions include low classification results, limited data set validation, a large number of channels, high classification energy, large area, and lack of a fully-integrated SoC solution.

ASD children face frequent and swift mood changes as compared to neurotypical or typically developing children. Past research has shown that these emotional disorders are not only key symptoms of ASD, but also associated with other core symptoms including repetitive behaviors [15]. It is observed that ASD patients have constricted valence and arousal scales as compared to typically developing children [15]. The Amygdala region of the brain is primarily responsible for emotion processing, and ASD patients suffer from an enlarged amygdala volume [81]. The aggression, anxiety, depression, and negative emotions in ASD result from the atypical amygdala region and are reflected by narrower valence and arousal scales [82].

The self-injurious emotions have reported higher levels of negative valence and high arousal [83]-[84]. A NEOB is therefore defined to represent the self-harming emotions with maximum valence negativity and maximum arousal [83], [84]. Each human emotion is predicted as self-harming or non-self-harming emotion. The self-harming and non-self-harming emotions are represented by emotions with NEOB and emotions without NEOB, respectively. It has been observed that the electrical onset in EEG signals of the NEOB occurs before the physical outcome [85]-[86].

The DEAP and SEED emotions prediction data sets are used for the NEOB prediction [24]. The positive and negative valence and arousal for NEOB prediction are defined as valence values between 7 to 9, and 1 to 3 respectively. A NEOB is defined as an emotion

with minimum (negative) valence and maximum (positive) arousal. For example in subject #1, the emotional trial numbers 31, 32, 34, and 35 are used to represent the NOEB in DEAP data set [24].

Since the SEED data set only provides valence labels [36]. The negative valence values were used to defined as NEOB. The emotional trials numbered 3, 4, 7, 12, and 15 correspond to NEOB in the SEED data set for all subjects. Figure 5-1 is used to portray an ASD mind with and without NEOB using valence and arousal scales.
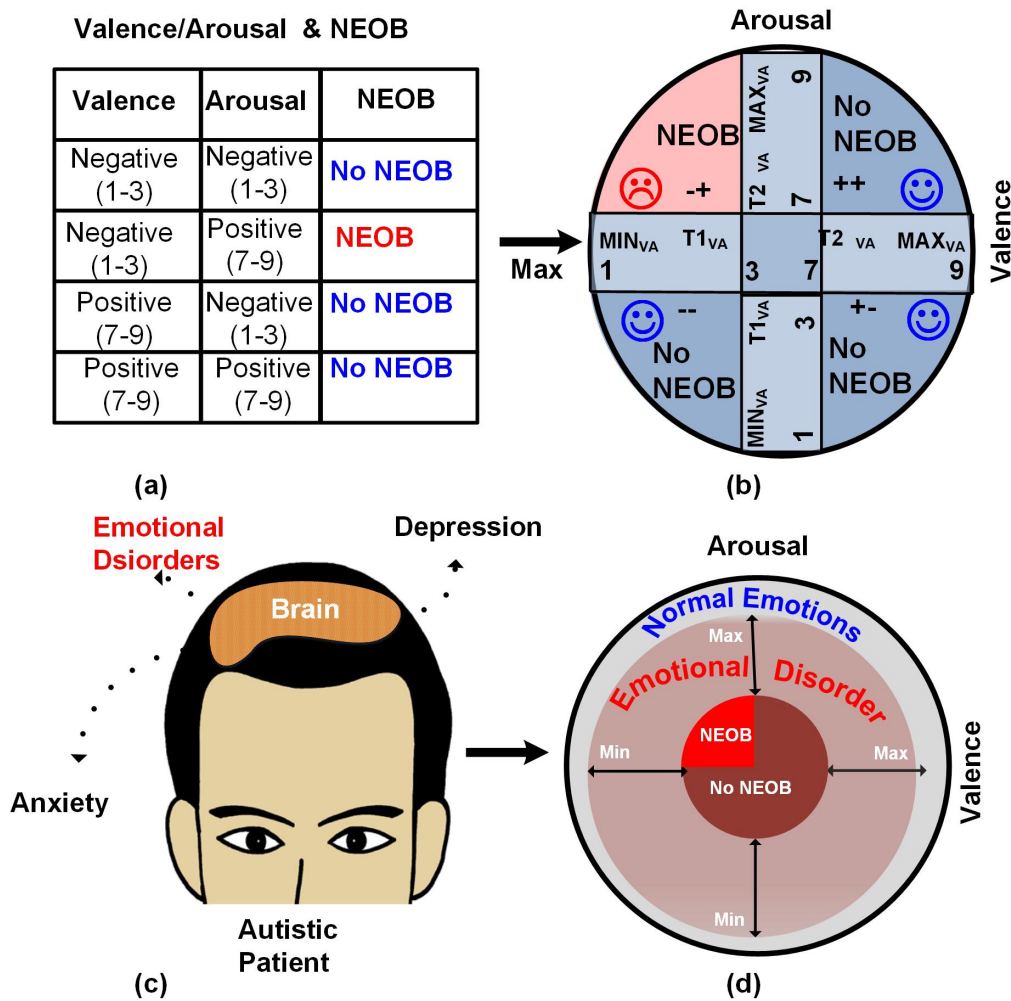
Figure 5-1: (a) Valence Arousal Classification & Corresponding Emotion, (b) Valence Arousal Thresholds & Minimum, Maximum Value (c) ASD Patient's Mind With Emotional Disorders (d) Constricted Valence Arousal Scale.

Figure 5-1 (a) and Figure 5-1 (b) show the procedure to mark an emotion with NEOB

or without NEOB (No NEOB) using DEAP data set and Russel's valence arousal scale [24], [34]. The positive and negative valence and arousal threshold values (1-3 and 7-9) are used to mark NEOB, as depicted in Figure 5-1 (b). Figure 5-1 (c) portrays the brain of an ASD patient for such emotional disorders dominated by negative emotions. The constricted valence and arousal scales for NOEB can be handled by altering the minimum and maximum valence and arousal scales as depicted in Figure 5-1 (d). The outer circle in Figure 5-1 (d) portrays the valence and arousal scales of a neurotypical person. The inner circles portray the constricted valence and arousal scales of an ASD patient. The change of threshold can be handled by simply uploading new classification parameters as discussed earlier in section 4.3.3 for DBE^1.

5.2   NEOB Prediction Processor Algorithm

Deep learning or DNN is a new dimension of ML, where the algorithms are inclined toward human brain structure. DNN has recently shown great performance over conventional ML algorithms for human emotions prediction [25]. The DNN eliminates the conventional feature extraction process and the raw EEG data is directly fed to the DNN for classification as discussed earlier in chapter 2 (Figure 2-6).

Despite the excellent results, the DNNs are mostly used for software-based systems and avoided in hardware implementations targeting emotion prediction using EEG signals. The few DNN-based hardware implementations have either used customized feature extraction with DNN based classifier or the features are extracted offline [25]-[26]. The offline feature extraction with the on-chip implementation for emotions prediction does not provide a fully on-chip wearable implementation. The fully on-chip implementation of a DNN for emotions prediction requires millions of classification parameters, and hence not feasible to provide an ULP and miniaturized SoC. I, therefore, targeted a customized DNN implementation after suitable feature extraction with a minimum number of channels. It was targeted to implement the DBE^2 NEOB prediction SoC with 2-4 low complexity features per EEG channel utilizing two to four EEG channels.

5.2.1 Limited Channel Selection

Previous research has proposed that the frontal lobe is highly related to negative emotions [87]. The frontal lobe EEG channels are highly suitable for wearable systems for real-time and continuous brain activity monitoring [88]. The frontal EEG electrodes are very convenient and suitable for dry electrode headsets [88]. The dry electrode headsets with frontal EEG channels can avoid the uncomfortable adjustment of electrodes through hair [88].

I, therefore, targeted to utilize a subset of two to four EEG channels from the set of frontal EEG channels (FP1, FP2, F3, F4, F7, F8, AF3, and AF4). F3 and F4 electrodes were identified to be the most discriminating for NEOB prediction among these eight electrodes. This channel combination is feasible for ASD patients due to minimal discomfort and unease [88]. They utilized 9.5X fewer resources (gate count) compared to our previous eight-channel pool [38], [72] using DNN. The selection of the most relevant and minimum number of EEG channels eases the job for the feature extraction and classification algorithms to provide a higher classification accuracy [39].

5.2.2 Feature Selection

The customized DNN-based NEOB prediction processor requires suitable features and DNN architecture selection for NOEB prediction. A separate and dedicated feature set is selected instead of conventional DNN implementation to avoid huge on-chip EEG data and parameters storage [39], [89]. The feature selection process involves the selection of suitable feature vectors which can utilize the real-time EEG signals for NEOB prediction with maximum classification results. I analyzed a set of different time and frequency domain features for the selected EEG channels. The experimented set of features includes mean, standard deviation, Kurtosis value, Hjorth parameters, Skewness (SKEW), ZCD, wavelet transforms (db4), and spectral energy in different frequency bands.

A set of two time-domain features including SKEW and ZCD was selected for the NOEB prediction after an extensive analysis between the classification results and the hardware implementation cost. The hardware implementation cost was calculated using the area (number of gates), and the power utilization of these features. The wavelet transform and the feature vector including ZCD and SKEW provided the best classification results. But

the hardware cost (area x power) of the wavelet transform was $> 39X$ higher than the feature set including ZCD and SKEW.

ZCD feature embeds useful information for emotions prediction [90]. SKEW is also one of the most important statistical features for NEOB prediction [91]. The ZCD measures the frequency of sign changes in the EEG signal. The SKEW represents the amplitude regularity of the EEG signal. Equations 5.1 and 5.1 are used to calculate the ZCD and SKEW respectively. The SKEW is calculated after the calculation of mean(μ) and standard deviation (σ) of the EEG signal. k is an integer value that varies from 1 to N in equations 5.1 and 5.2. N is the total number of EEG samples. EEG_k and EEG_{k-1} are used to represent the current and preceding EEG samples, respectively. μ_{EEG} and σ_{EEG} are used to represent the mean and standard deviation of the EEG signal respectively.

$$ZCD = \left[\sum_{k=0}^{N} [Sign(EEG[k]) \neq Sign(EEG[k-1]) \mid EEG[k] \neq 0 \right] \quad (5.1)$$

$$SKEW = \left[\sum_{k=0}^{N} [\frac{(EEG[k] - \mu_{EEG})^3}{(N-1) * \sigma_{EEG}^3} \right] \quad (5.2)$$

Since, a subset of two EEG channels including F3 and F4 channels was selected for this DBE^2 NEOB prediction processor. The selected feature vector (FV) contains a set of four features. The selected FV includes $ZCD_{F3}, ZCD_{F4}, SKEW_{F3}$, and $SKEW_{F4}$. ZCD_{F3}, and ZCD_{F4} represent the number of zero crossings for F3 and F4 channels, respectively. $SKEW_{F3}$, and $SKEW_{F4}$ represents the skeweness for F3 and F4 channels, respectively. Equation 5.3 is used to represent the selected FV.

$$FV = \begin{bmatrix} ZCD_{F3} \\ ZCD_{F4} \\ SKEW_{F3} \\ SKEW_{F4} \end{bmatrix} \quad (5.3)$$

The ZCD and SKEW features have quite different range of values. The ZCD feature

ranges in thousands whereas the SKEW feature ranges in fractions. Therefore, the selected FV is then normalized to a uniform range using Z-score normalization. The normalized FV (FV_{Norm}) is represented in equation 5.4 which is calculated after Z-score normalization of FV. A detailed analysis of these features was carried out for NEOB prediction.

$$FV_{Norm} = \left[ZScore(FV) \right] \tag{5.4}$$

Figure 5-2 shows the center scattered plot for ZCD and SKEW features for emotions corresponding to NEOB and without NEOB for 12 different randomly chosen trails in DEAP data set. It can be observed that their is a clear separation between the ZCD for NEOB and emotions without NOEB as shown in Figure 5-2 (a). The emotions corresponding to NEOB and are mostly on one side of the center, whereas the NO NEOB's are on both sides of the center. Similarly, the emotions corresponding to NEOB and without NEOB have a clear separation for SKEW feature as shown in Figure 5-2 (b). The emotions corresponding to NEOB and are mostly on one side of the center, whereas the NO NEOB's are on other side of the center.

The customized DNN classifier model used in this work benefits from this clear separation between positive and negative classes to achieve the high accuracy for NOEB prediction (85.4%) for DBE^2 NOEB prediction processor. The other hidden information is learned by the proposed DNN classifier for NOEB prediction.

5.2.3   Deep Neural Network Classifier

A customized feed-forward DNN is used for the NEOB prediction using the selected FV. The DNN classifier contains an input layer (IL), an output layer (OL), and two hidden layers. The hidden layers are added to handle the non-linearity of data for classification, and more complex classification problems require more hidden layers [92]. However, the hardware complexity for the on-chip implementation of DNN classifier increases with the addition of each hidden layer (HL). The area and classification energy for the hardware implementation of DNN is directly affected by the number of hidden layers. Therefore, it was aimed to utilize a DNN with a minimum number of hidden layers.

The proposed fully connected DNN classifier contains two hidden layers (HL1 and HL2),
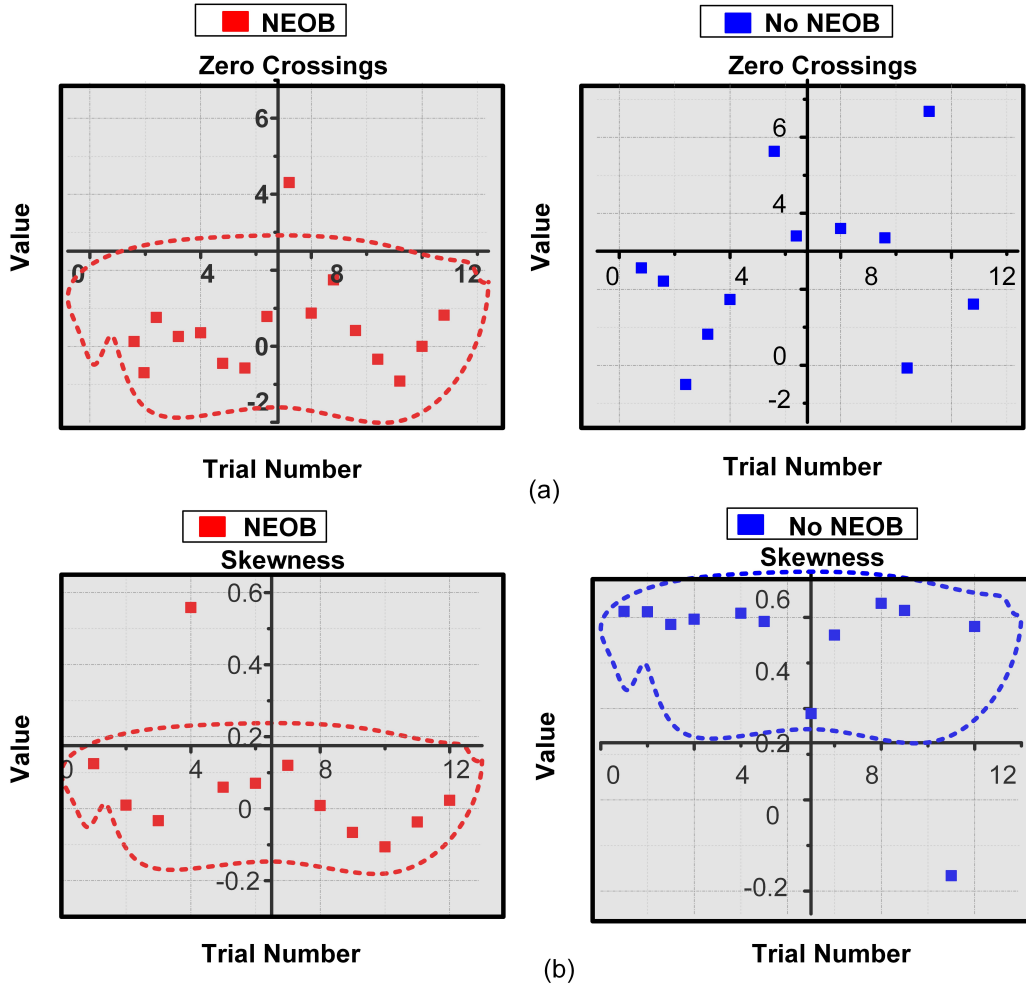
Figure 5-2: Z-Score Normalized ZCD & SKEW for Different (a) NEOB's & (b) Without NEOB Emotions.

an IL and, an OL. The IL, HL1, HL2, and OL contain 8, 16, 32, and 1 node with a structure of 4 x 8, 8 x 16, 16 x 32, and 32 x 1, respectively. Figure 5-3 shows the architecture of the fully connected DNN for NEOB prediction.

The activation functions in a DNN are used to activate or deactivate certain neurons in a DNN [93]. A sigmoid (SGM) activation function provides fast learning and is suitable for binary classification problems [94]. The RLU activation function speeds up the learning and classification process by nullifying certain neurons [94]. The linear, SGM, RLU, and SGM activation functions were used for the IL, HL1, HL2, and OL, respectively. The proposed fully connected DNN was trained for 1000 epochs to calculate the best weights and biases for NOEB prediction.

Fully Connected Deep Neural Network Classifier
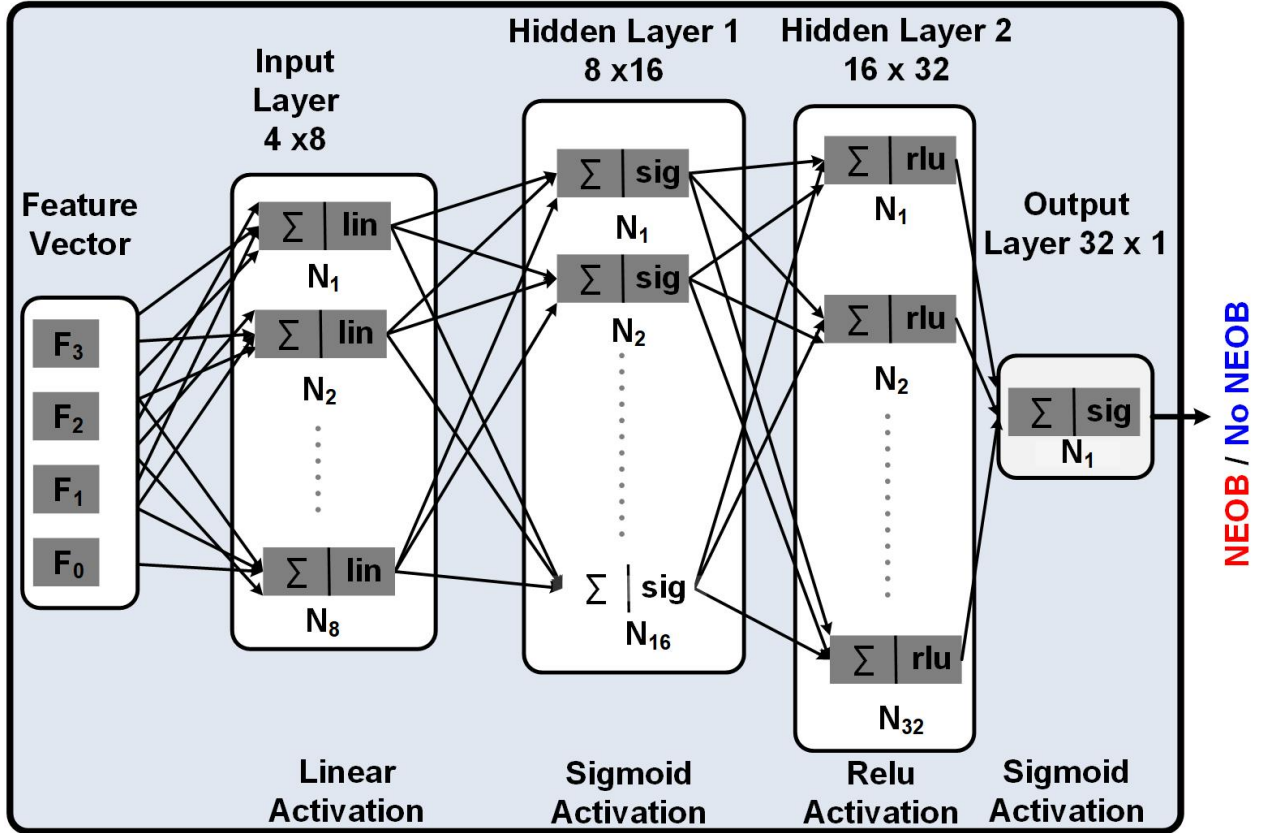
Figure 5-3: Fully Connected DNN Classifier Architecture for DBE^2.

The Z score normalized FV was forwarded to the DNN for NEOB prediction. Equation 5.5-5.9 are used for NEOB prediction using the selected features. NFV3-NFV0 represents the normalized FV in equation 5.5. The IL multiplies and accumulates the normalized features with the IL weights $(W0_0 - W0_7)$ and biases $(b0_0 - b0_7)$, as in equation 5.5. The output for IL is forwarded as $Z0_0 - Z0_7$ to the first HL (HL1).

$$
[\ Z0_7\ ...\ Z0_0\ ] = [NFV3....NFV0].\begin{bmatrix} W0_3 & W0_7 & .... & W0_{31} \\ W0_2 & W0_6 & ... & W0_{30} \\ W0_1 & W0_5 & ... & W0_{29} \\ W0_0 & W0_4 & ... & W0_{28} \end{bmatrix} + \begin{bmatrix} b0_7 & .... & b0_0 \end{bmatrix}
$$

$$(5.5)$$

$$[Z1_{15}...Z1_0] = SGM\left([Z0_7 \ Z0_0] \cdot \begin{bmatrix} W1_7 & W1_{15} & .... & W1_{127} \\ W1_6 & W1_{14} & ... & W1_{126} \\ . & . & & . \\ W1_0 & W1_8 & ... & W1_{120} \end{bmatrix} + \begin{bmatrix} b1_{15} & .... & b1_0 \end{bmatrix}\right)$$

$$(5.6)$$

The HL1 multiplies and accumulates the IL output $(Z0_0 - Z0_7)$ with the HL1 weights $(W1_0 - W1_31)$ and biases $(b1_0 - b1_{15})$, as in equation 5.6. The output for HL1 is forwarded as $Z1_0 - Z1_{15}$ to the second HL (HL2).

$$[Z2_{31} \ ... \ Z2_0] = RELU\left([Z1_{15} \ Z1_0] \cdot \begin{bmatrix} W2_{15} & W2_{31} & .... & W2_{511} \\ W2_{14} & W2_{30} & ... & W2_{510} \\ . & . & & . \\ W2_0 & W2_{16} & ... & W2_{496} \end{bmatrix} + \begin{bmatrix} b2_{31} & .... & b2_0 \end{bmatrix}\right)$$

$$(5.7)$$

The HL2 multiplies and accumulates the HL1 output $(Z1_0 - Z1_{15})$ with the HL2 weights $(W2_0 - W2_{511})$ and biases $(b2_0 - b2_{31})$, as in equation 5.7. The output for HL2 is forwarded as $Z2_0 - Z2_{31}$ to the OL.

$$[Z3] = SGM\left(\begin{bmatrix} Z2_{31} & Z2_{30} & . & Z2_0 \end{bmatrix} \cdot \begin{bmatrix} W3_{31} \\ W3_{30} \\ . \\ W3_0 \end{bmatrix} + \begin{bmatrix} b3 \end{bmatrix}\right) \quad (5.8)$$

The OL multiplies and accumulates the HL2 output $(Z2_0 - Z2_{31})$ with the OL weights $(W3_0 - W3_{31})$, and bias (b3), as in equation 5.8. The output for OL is forwarded as Z3 for the final emotion prediction.

$$NEOB = \begin{bmatrix} 1 \,,\, if \quad Z3 < 0.5 \\ 0 \,,\, if \quad Z3 \geq 0.5 \end{bmatrix} \tag{5.9}$$

Since, the decision boundary of the sigmoid activation function is centered around 0.5 for positive and negative classes separation. Therefore, a NEOB is predicted by comparing the output for OL with 0.5 as in equation 5.9.

The complete NEOB classification process using limited (two) EEG channels with separate feature extraction and DNN classification is depicted in Figure 5-4. The EEG signals for F3 and F4 channels are forwarded for feature extraction and DNN classification using a fully connected DNN. The EEG data of two channels for 60 seconds duration is forwarded to the feature extraction engine. The DNN-based NEOB classification processor is trained on the EEG data sampled on a sampling frequency of 128 Hz, which requires 7680 samples for an EEG segment of 60 seconds duration per channel. The two-channel DNN-based NEOB classification processor requires 15360 EEG samples to calculate the ZCD and SKEW features. A total number of 15500 clock cycles are required to establish the features including the clock cycles required to sample the input EEG signals for the two channels and the multiplexing operations for the twin channel EEG operation. The features are established in 15.14 seconds with the 1 kHz clock provided to the feature extraction engine. Figure 5-4 shows the selected segments of the EEG signal to depict the NEOB classification procedure. The ZCD calculates the zero crossings of the EEG signal. The EEG signal is marked as red at the time instances on which it changes its sign. The SKEW calculates the amplitude regularity of the signal through normalized EEG using mean and standard deviation.

5.2.4 NEOB Prediction System Architecture

The SoC block diagram of the proposed DNN-based patient-specific NEOB classification processor is shown in Figure 5-5 . The AFE is comprised of two channels with a twin-channel sharing capacitive sampling (TCSCS) technique that contains a parallel on resistance channel sharing multiplexer (PRONCSM) to switch between the two channels [39]. The channel switching frequency (CSF) of 500Hz multiplex both the capacitive coupled low noise amplifier (C2LNA) and continuous-time digitally assisted capacitive coupled (CTDC2) instrumentation amplifier. To maximize the common mode rejection ratio (CMRR), all ca-
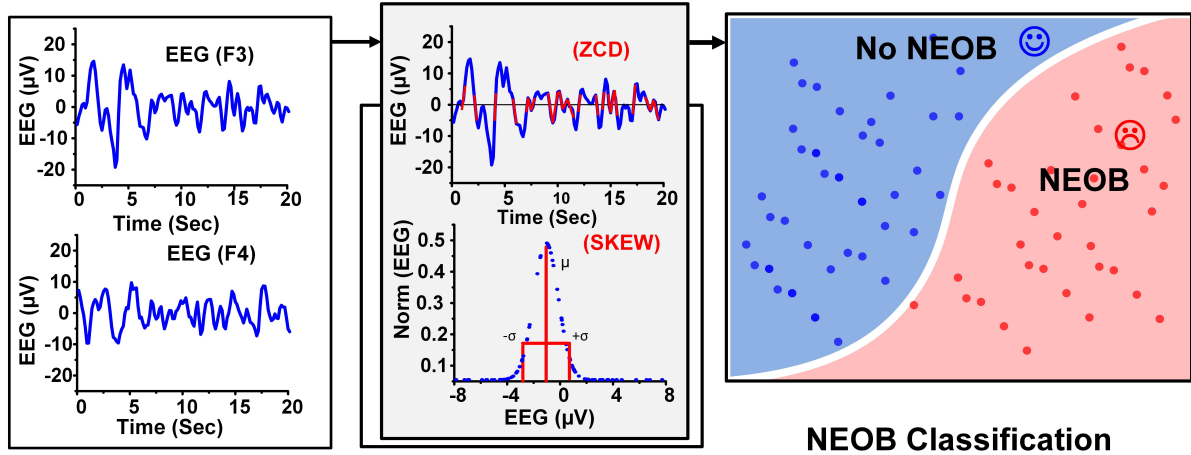
Figure 5-4: Negative Emotion Outburst Prediction Using an EEG Signal
Highlighting Features & Deep Neural Network Classification

pacitors used in C2LNA and CTDC2 are integrated on-chip to ensure maximum matching. The amplified and filtered EEG data is digitized using a 10-bit successive approximation analog-to-digital converter (SAR-ADC) [39]. The digitized data is processed in the DNN-based patient-specific classification processor for NEOB prediction. The AFE is integrated with the proposed DBE processor for NOEB prediction (DBE^2) in this Ph.D. project. The SoC implementation and the hardware architecture of DBE^2 are explained in the next section.

5.2.5 NEOB Prediction DBE Processor Implementation

The on-chip emotion prediction FEE and classification are done in the DNN based patient-specific classification processor. Figure 5-5 shows the architecture of the DNN processor, which incorporates a FEE, and a DNN classification engine (DNN-CE) for the NEOB or four-state emotion classification, followed by an emotion decision logic (EDL) block. The classification engine can be used for the NEOB classification or the four-state emotion classification, i.e., (happy, sad, relaxed, and angry) using standard biomarkers for emotion classification (valence and arousal). The valence and arousal classification parameters are uploaded to the DNN classification engine for the four-state emotion classification. The classification parameters upload is performed in a similar way as DBE^1, explained earlier in Figure 4-13, Section 4.3.3. The 2x1 multiplexer (MUX) and the EDL are used to select the
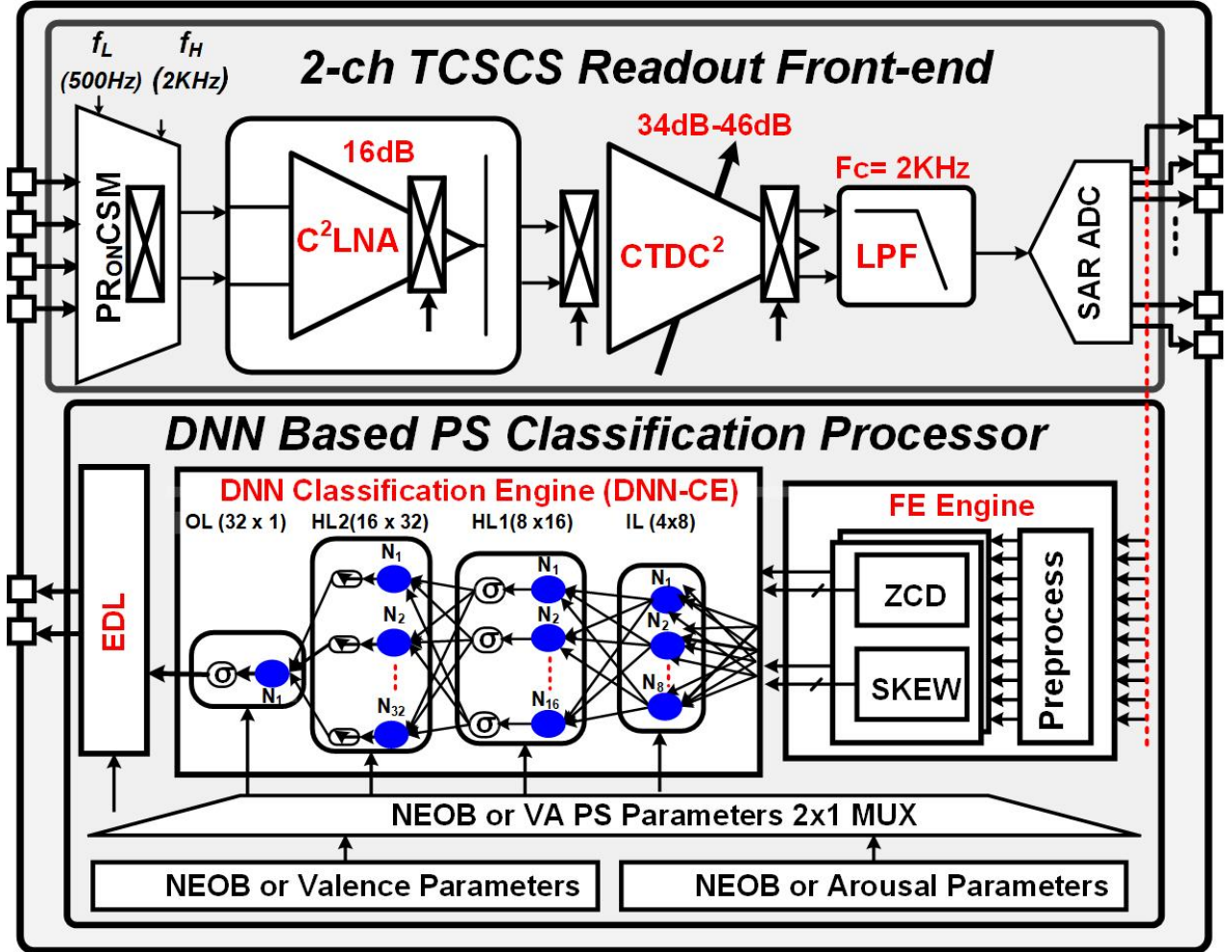
Figure 5-5: Deep Neural Network Negative Emotion Outburst Prediction Processor Architecture.

valence arousal (VA) or NEOB classification parameters and, perform NEOB prediction as depicted in Figure 5-5. The AFE of [39] was integrated with the DBE^2 processor designed in this Ph.D. work.

This SoC provides the NEOB classification closed loop biofeedback to the caregiver in real-time with < 1 min latency during a learning or cognitive skill activity. This continuous biofeedback can avoid NEOB and consecutive self-injuries.

Feature Extraction Engine Implementation

The FEE extracts ZCD and SKEW features from the incoming EEG signal. Figure 5-6 shows the hardware architecture of the ZCD implementation unit. The 10-bit incoming

quantized EEG signal is forwarded to the ZCD unit after preprocessing to remove noise and other artifacts. A customized comparator (comp1) determines that the incoming EEG signal is zero, positive, or negative and raises the flags A, B, or C accordingly. A state table (6b x 3b) compares the current state along with these flags and provides the next state using a state register (S Reg). Two different states (S3 and S5) are used to detect the positive-to-negative or negative-to-positive sign changes using comparators comp1 and comp2, respectively. An OR gate increments a counter (Count) using an increment (INC) signal, if there is any sign change. The ZCD for the two channels is stored in a memory block (2 x 32 bits) as ZC_1 and ZC_0. The ZC_1 corresponding to channel 1 (CH_1) and ZC_0 corresponding to channel 2 (CH_2) are then and forwarded to the DNN classification unit as FN_3 and FN_2, respectively, after feature normalization.
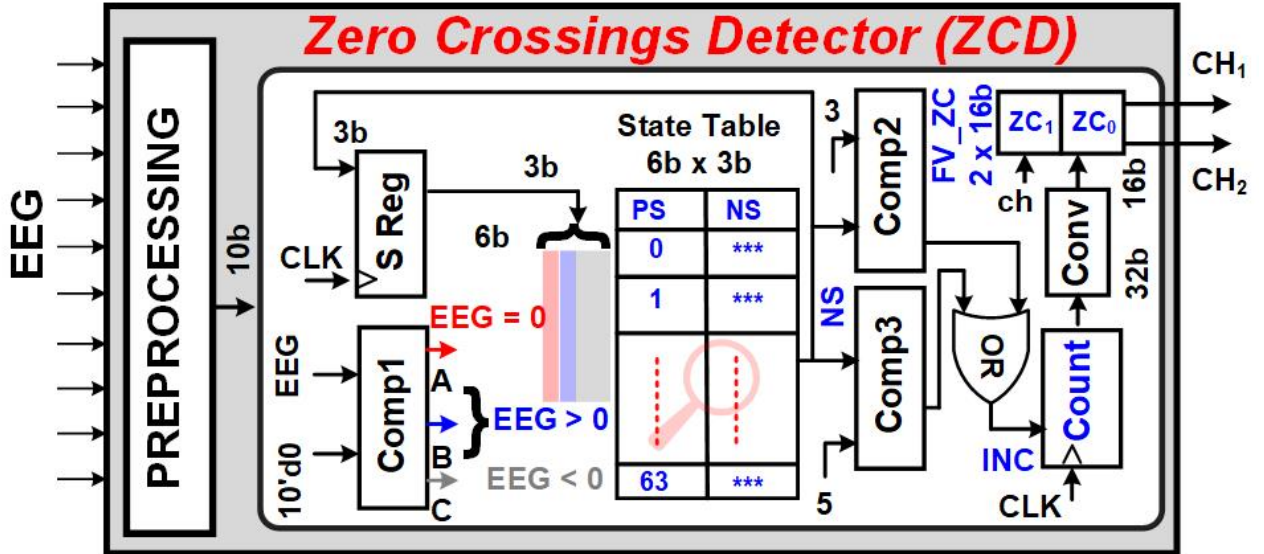
Figure 5-6: Hardware Architecture for ZCD Implementation.

The conventional SKEW implementation involves the calculation of the cubic difference of μ from each sample to calculate $(EEG[k] - \mu)$, $(EEG[k] - \mu)^2$, and $(EEG[k] - \mu)^3$ as in equation 5.2. It also requires the calculation of cubic standard deviation (σ^3). These calculations require the storage of complete EEG time series requiring a huge memory of $\approx$ 10 MB for a 2-channel operation.

Since we are performing a binary classification, the exact values of SKEW are insignificant. We exploited this fact and mathematically analyzed the SKEW equation after ignoring EEG[k] in equation (5.2), which results in an ASKI as mentioned in equation 5.10. The

revised ASKI equation is further simplified as equation 5.11-5.13 to propose equation 5.13 for calculating ASKI. The value of N in equation 5.13 is data set dependent. It depends on the sampling frequency and the time duration for which a NEOB classification is performed from the EEG signals.

$$ASKI = \left[\sum_{k=0}^{N}\left[\frac{-\mu_{EEG}^3}{(N-1)*\sigma_{EEG}^3}\right]\right] \tag{5.10}$$

$$ASKI = \left[\sum_{k=0}^{N}\left[\frac{-\mu_{EEG}^3}{(N-1)*sqrt(\frac{\sum_{\mu_{EEG}^2}}{N})}\right]\right] \tag{5.11}$$

$$ASKI = \left[\sum_{k=0}^{N}\left[\frac{-\mu_{EEG}^3}{(N-1)*\frac{\sum_{\mu_{EEG}}}{sqrt(N)}}\right]\right] \tag{5.12}$$

$$ASKI = \left[\sum_{k=0}^{N}\left[\frac{EEG[k]^3}{(N-1)*sqrt(N)*N^3}\right]\right] \tag{5.13}$$

The variation of N does not cause any significant impact on the classification metrics of the system if the selected duration of EEG signals for NOEB prediction is ≥ 45 seconds. The analysis of the ASKI trend in comparison with SKEW for binary classification of NEOB for 10 random trials was performed. It depicts that the ASKI contains some partial information about SKEW. The rest of the information for NEOB prediction is learned by the DNN classifier.

Figure 5-7 shows the hardware architecture of the proposed ASKI (equation 5.13) in comparison with the conventional SKEW implementation (equation 5.2). The hardware architecture of the proposed ASKI is shown in Figure 5-7 (b). The ASKI only requires the prior calculation of the sum of EEG signals. The mean (μ) in equation 5.11-5.12 is calculated after summing up the EEG signal and then dividing (inverse multiplication) with the total number of samples, as simplified in equation 5.13. Similarly, the (μ^3) is calculated after the multiplication of (μ) and (μ^2), as simplified in equation 5.13. The division operations in
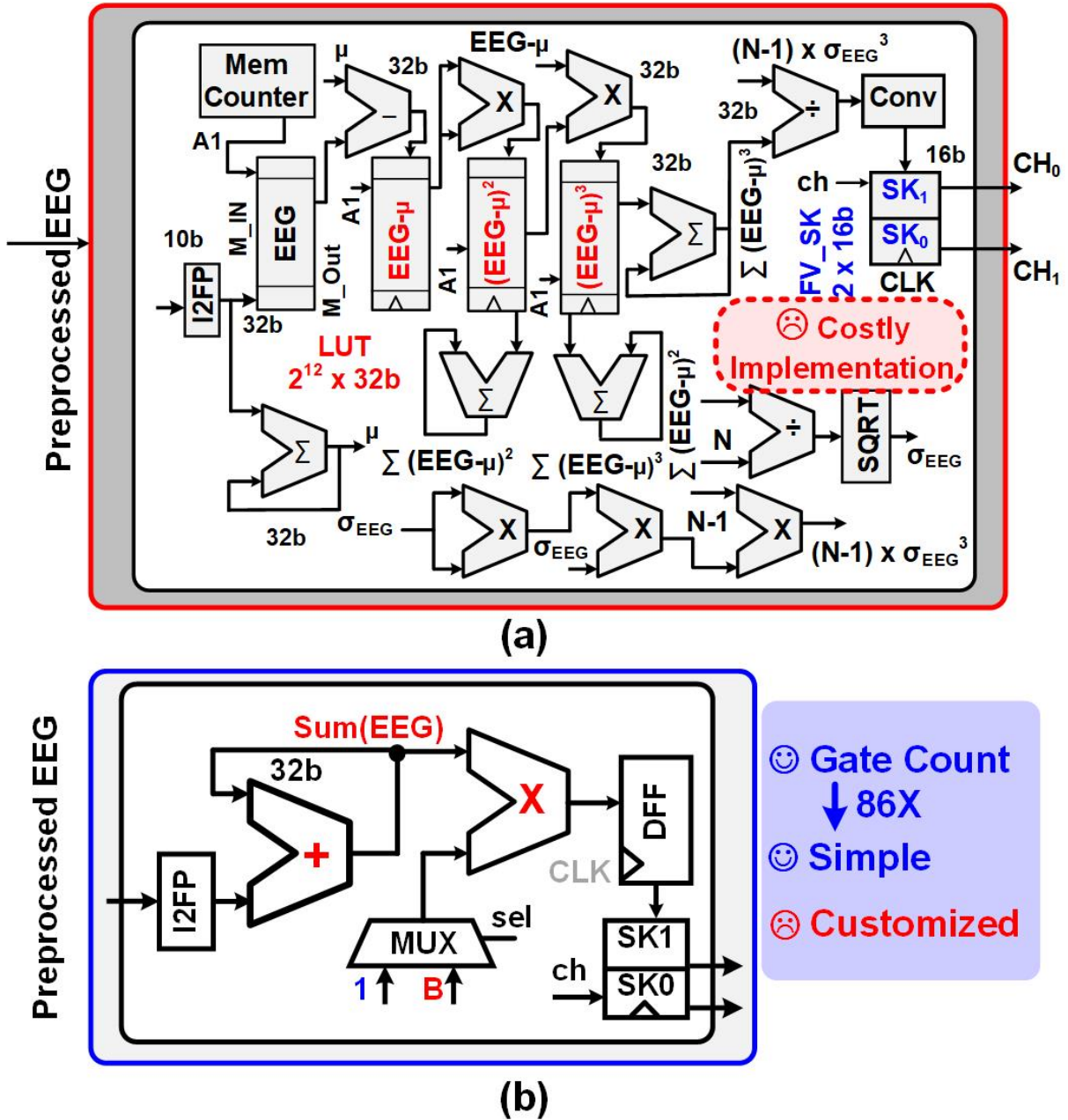
Figure 5-7: Hardware architecture for (a) Conventional SKEW implementation (b) Proposed ASKI implementation.

ASKI can simply be replaced by inverse multiplications. The ASKI for the selected channels is stored in a memory block (2 x 32 bits) as $(SK_0 \; and \; SK_1)$.

The conventional SKEW implementation, however, requires the storage of a large number of EEG samples to calculate (μ) , and then (μ^2), and (μ^3), as shown in Figure 5-7 (a). This ASKI implementation resulted in gate reduction by 86X as compared to the costly

implementation of conventional SKEW, without compromising the classification results. The (SK_0 and SK_1) calculated by the ASKI unit are forwarded to the classification unit as (FN_0 and FN_1) respectively after feature normalization. The FN_0, and FN_1 correspond to the two selected EEG channels CH_0, and CH_1, respectively.

Deep Neural Network Classification Unit Implementation

The DNN classification unit implements the proposed four-layer fully connected deep neural network classifier. The hardware implementation of the fully connected DNN classifier unit requires the multiplication and accumulation (MAC), SGM, and RELU (RLU) calculation units. The MAC operations involve the multiplication and addition operations for each layer using classification parameters (weights and biases).

Figure 5-8 shows the hardware architecture of the proposed semi-pipelined fully-connected DNN classification unit for NEOB classification. A semi-pipelined arithmetic and logic unit (ALU) controlled by a finite state machine design is used to control the ALU operations. The ALU operations are controlled by the opcode (ctrl) and the input operands provided by the fully connected DNN control unit. The input operands are either weights, biases, or intermediate results based on the DNN architecture. The ALU performs the addition, multiplication, SGM, RLU, or no operation (NOP) instructions on each clock cycle. The NOP operations are inserted to avoid data hazards and resolve any data dependencies. The proposed methodology requires $\approx$ 34X lesser resources (gate count) than the conventional approach [95].

The fully connected DNN control unit directs the ALU about the number of multiplication and additions in the MAC operation, and the number of intermediate outputs to be stored in the memory block (MEM). The ALU contains four floating point multipliers (M3-0), four floating point adders (A3-0), one memory block (32 x 16 bits), one RLU, and one SGM (sg) unit.

Figure 5-9 highlights the fully connected DNN operations for IL at two different time instances during the classification process. The ALU performs the MUL operations (equation 5.5) at T = 1 to multiply the FV ($FN_3 - FN_0$) with the first column of weights ($W_3 - W_0$) using the four floating point multipliers ($M_3 - M_0$). The remaining weights ($W_7 - W_4$) to ($W_{31} - W_{28}$) are multiplied with the FV ($FN_3 - FN_0$) from T = 2 to T = 8. The addition operations are initiated at T = 2 to complete the MAC operations. The operations for IL

91
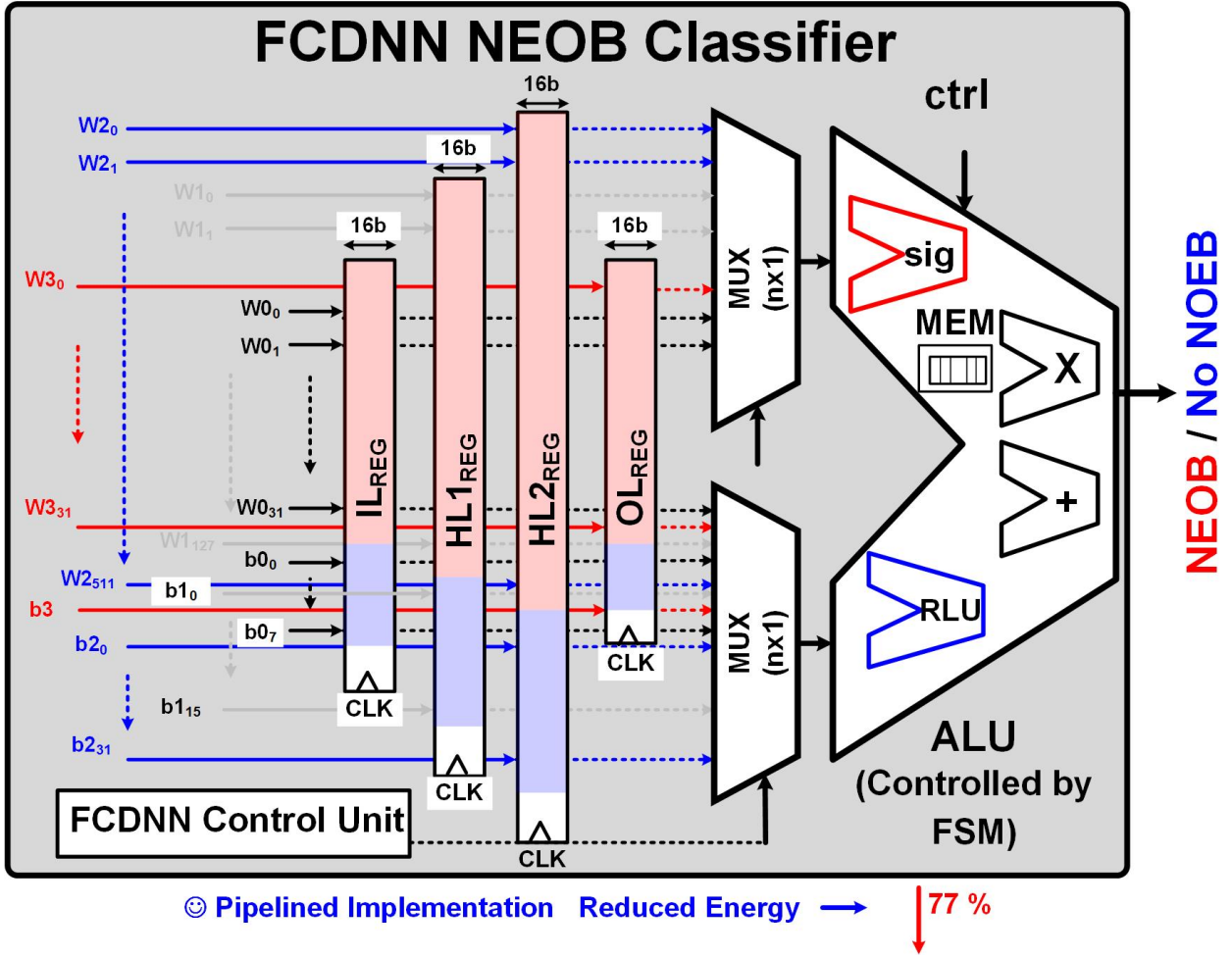
Figure 5-8: Proposed Fully Connected Deep Neural Network Architecture Hardware Implementation.

are completed at T = 21 to provide $Z0_3 - Z0_0$. The $Z0_3 - Z0_0$ is the output for IL, which is used as an input for the HL1.

Figure 5-10 shows the timeline for the semi-pipelined ALU-based fully connected DNN classification unit. The classification process is imitated by a low to high pulse (CF) after the calculation of FV. The IL, HL1, HL2, and OL operations are performed from T = 1 to T = 21, T = 22 to T = 190, T = 191 to T = 945, and T = 945 to T = 998, respectively.

$M_3 - M_0$ represents the outputs for four floating point multipliers for multiplication, and $A_3 - A_0$ represents the outputs for four floating point adders for addition. $G_{X,Y}$ is used to represent the output for the floating point adders. The X represents the output number associated with the addition for MAC operation and Y represents the number of additions.
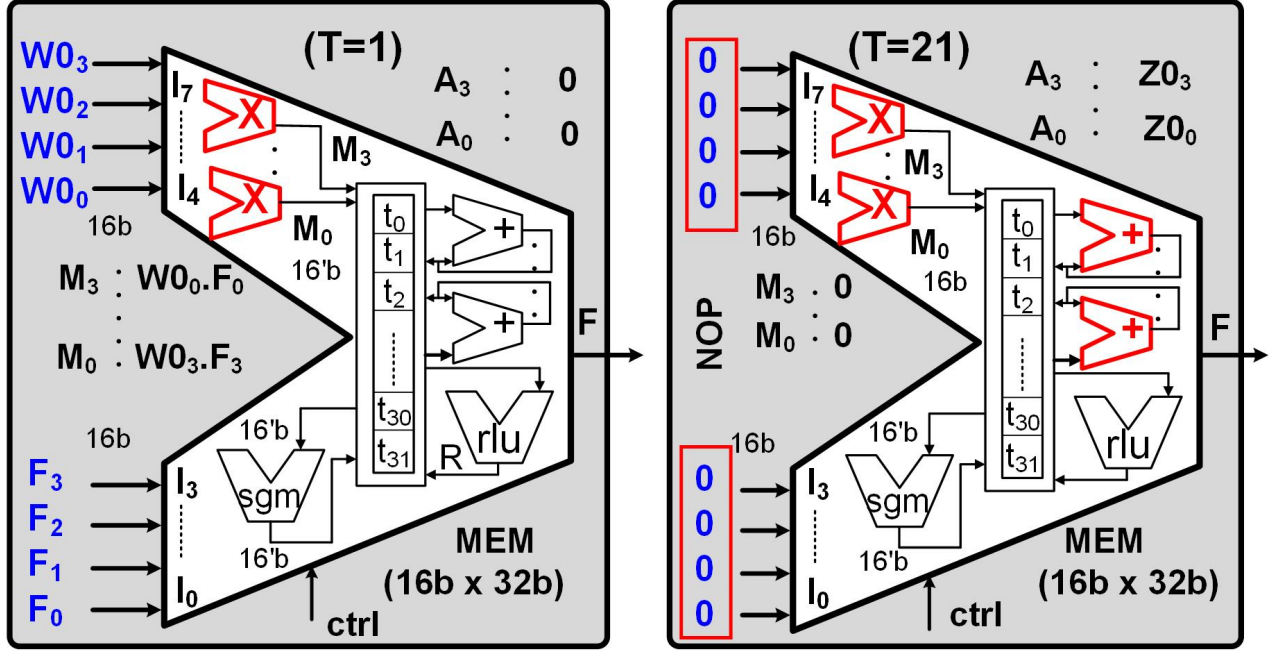
Figure 5-9: Fully Connected Deep Neural Network Classifier at Different Time Instances.

The FLP adder $(A_3 - A_0)$ performs the accumulation operations for $Z0_7 - Z0_4$ at T $= 2$ to T $= 9$.

$G_{7,1}$ at T $= 2$ represents that the current addition operation is associated with IL and the first addition operation is being performed. Similarly, $G_{7,2}$ to $G_{7,4}$ represent the second to fourth additions required for IL. Four, eight, sixteen, and thirty-two addition operations are required in each MAC operation for the IL, HL1, HL2 and, OL to calculate $Z0_7 - Z0_0$, $Z1_{15} - Z1_0$, $Z2_{31} - Z2_0$ and Z3, respectively. A 16 x 32 bits memory block is used to store the intermediate outputs. The proposed semi-pipelined implementation reduced the classification energy by 77% compared to the conventional fully connected DNN architecture [95].

The hardware cost of a SoC based processor is primarily dependent on its area. The overall area of the proposed DNN classification processor primarily depends on the number of FLP addition units, FLP multiplication units, activation function units, and the bit precision of these units. The bit precision of these units also effects the memory requirement for the classification parameters uploaded to the DNN classification processor. A fully parallel implementation of the proposed DNN classification processor with full precision
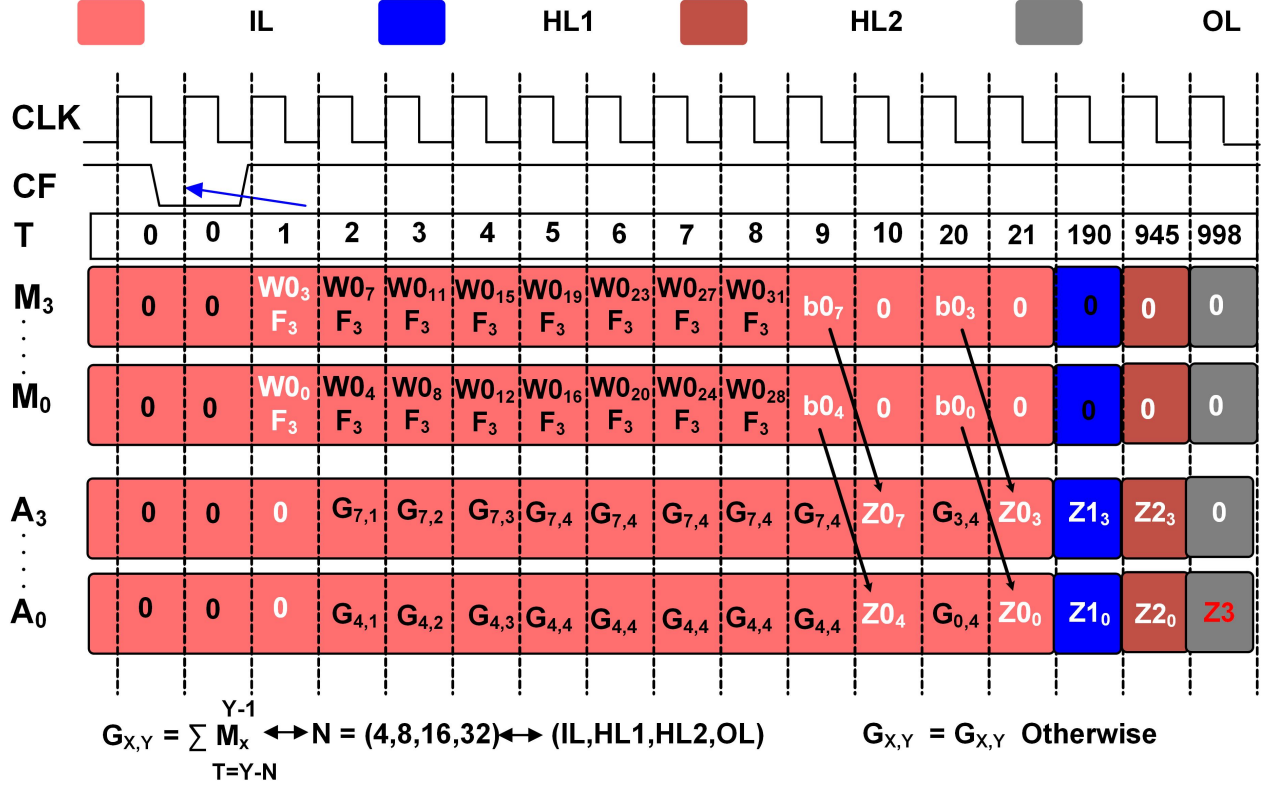
Figure 5-10: Timeline for Semi-Pipe-Lined Arithmetic and Logic Unit for Deep Neural Network classifier.

FLP numbers required a huge area ($> 100mm^2$). Therefore, I implemented the DNN processor in the semi-pipe-lined approach explained above, which required an overall area of $16mm^2$.

The RLU and SGM activation functions for the DNN classification are defined in equations 5.14 and 5.15 respectively. The RLU function is linear for positive values and the output of these values follows the input. However, it nullifies the negative values to zero as shown in equation 5.15. The hardware implementation for RLU is quite simple and just requires a two-to-one multiplexer.

$$RLU(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \qquad (5.14)$$

94

$$SGM(x) = \frac{1}{(1 + e^{-}x)} \tag{5.15}$$
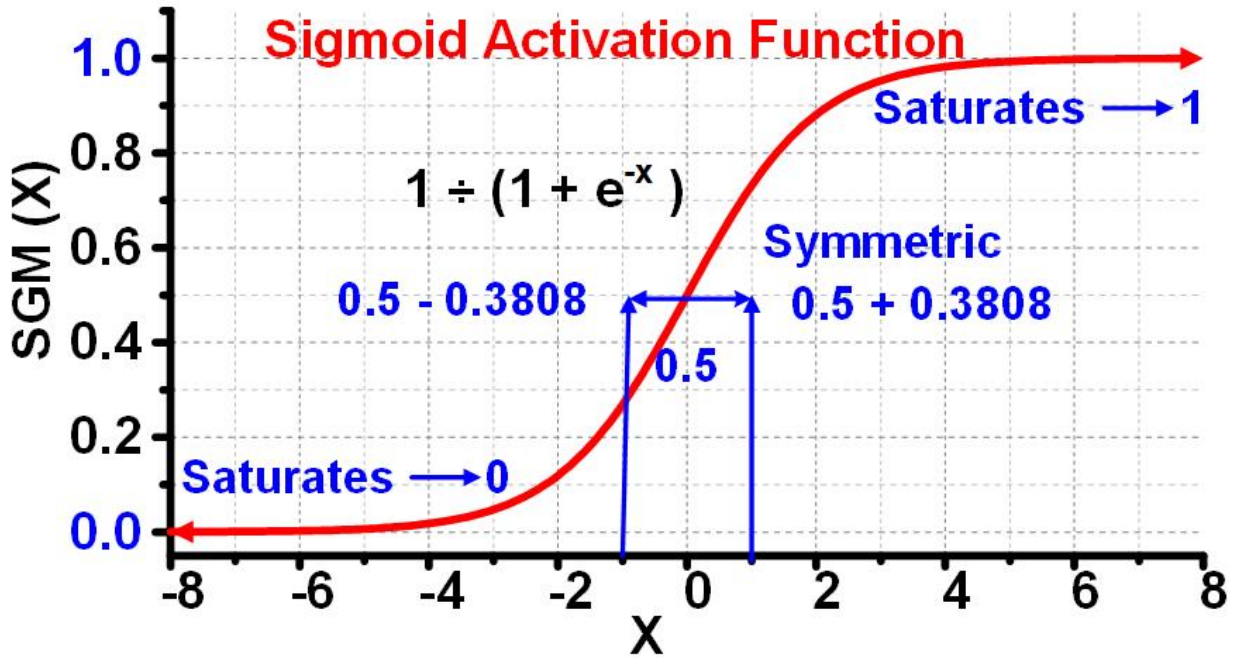
Figure 5-11: Sigmoid Activation Function Graph.

The SGM activation function predicts the output as a probability function between 0 and 1. The negative and positive inputs are mapped between 0 to 0.5, and 0.5 to 1, respectively. It has a complex hardware implementation as it involves the floating point exponential function and division operation as mentioned in equation 5.15. The SGM function is conventionally implemented by LUT, piecewise approximation, or CORDIC algorithm. The output values of the SGM function saturate near +/-8 with positive and negative symmetry ($x > 0$ and $x < 0$) to form an S-shaped curve as shown in Figure 5-11. This symmetry allowed us for a 50% lower memory usage for a LUT-based hardware implementation.

The hardware architecture for the SGM function implementation is shown in Figure 5-12. Since the SGM function inputs have floating point values, we scaled the values by 512 to shift them to an integer range. The scaling is performed by adding 9 to the mantissa values (X[14]-X[10]) of the input X. The scaled value is converted to integer format using a
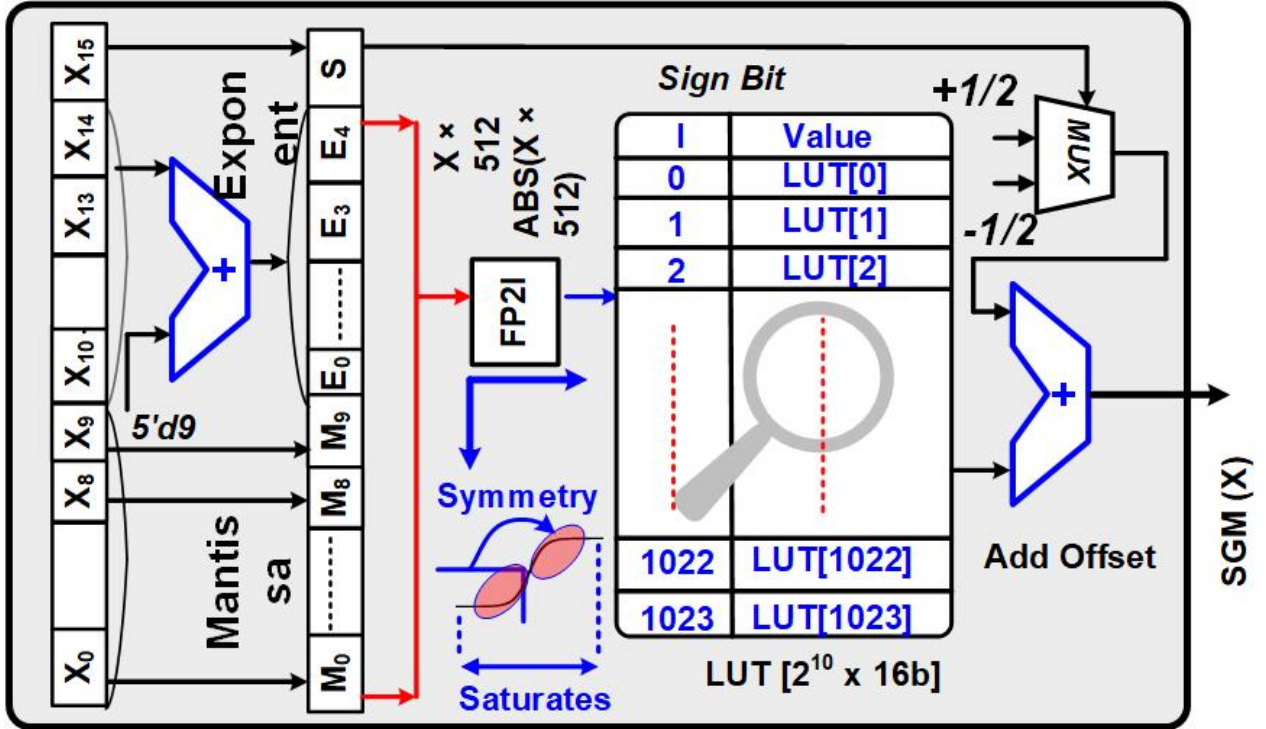
Figure 5-12: Sigmoid Activation Function Hardware Implementation.

floating-point to integer converter (FP2I). The integer index is then forwarded to a 2^{10} x 16 bits LUT. The LUT provides a shifted SGM value which is added to + or -0.5 based on the sign bit (X[15]) of the input X to obtain the SGM value. This hardware implementation avoids the expensive floating point implementation, which requires 512X more memory storage. The hardware architecture for the SGM activation function is shown in Figure 5-12.

The generic utilization of the proposed SGM unit cannot be ensured for all applications. Since the proposed SGM unit was implemented using a LUT-based approach. The LUT stores a total of 1024 SGM values as half-precision FLP numbers. The utilization of the SGM unit is dependent on the statistics of the input values, and the deviation of the input values from the values stored in the LUT. For e.g. the proposed SGM unit stores the 1024 shifted SGM values of 0.0068359375,0.013671875,...,7 in half-precision FLP format. Figure 5-13 shows the actual sigmoid values in full precision FLP format against the SGM values provided by the proposed SGM unit. It can be observed that the proposed SGM unit approximates the SGM values because of limited LUT size and half-precision FLP values. The SGM values provided by the proposed worked accurately for the emotions prediction

using DEAP and SEED data sets. But the generic utilization of the SGM unit is dependent on the DNN architecture, data statistics, and classification weights.
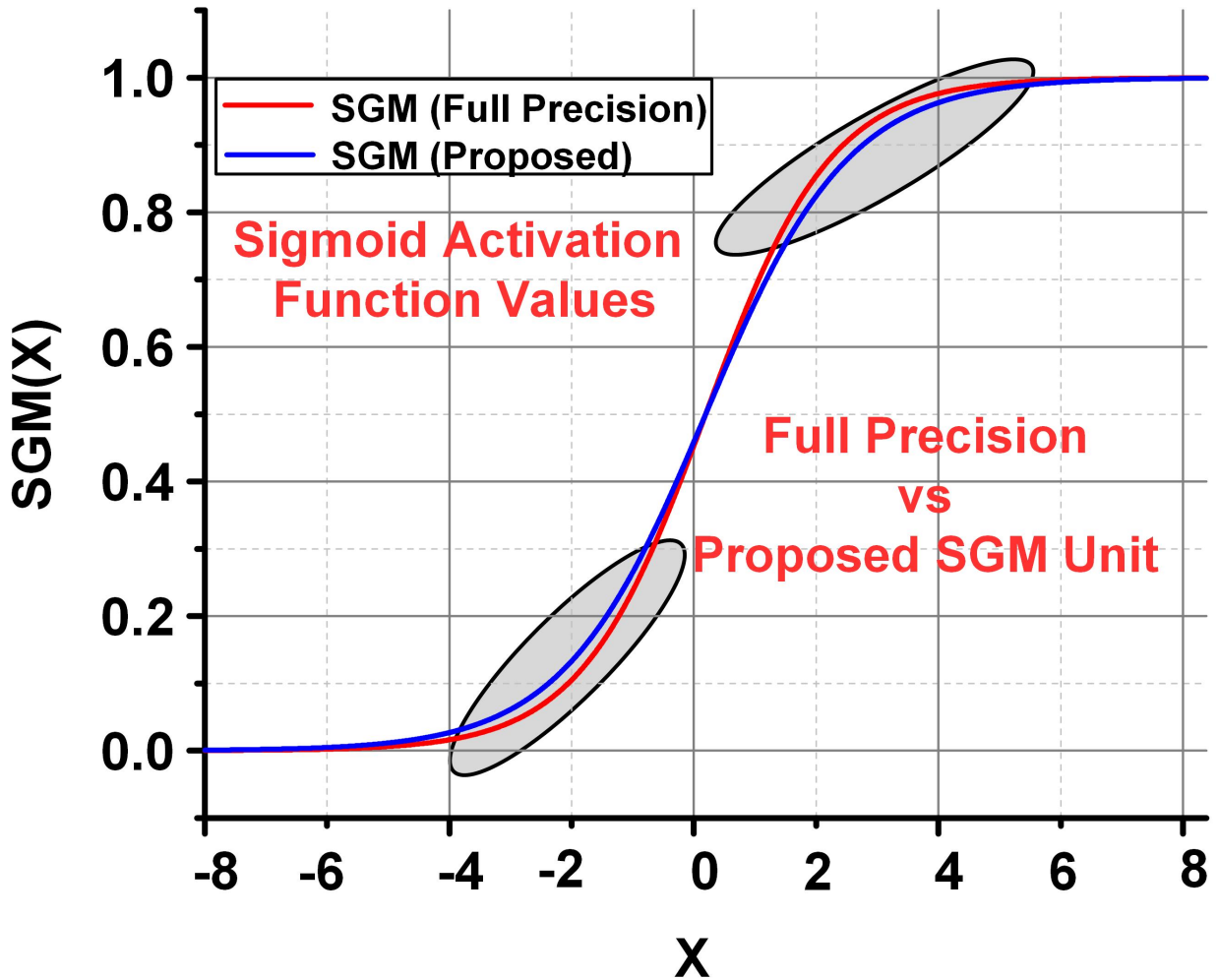
Figure 5-13: Sigmoid Values by Proposed Unit vs Sigmoid Values with Full Precision.

The ALU size (number of bits) is an important parameter for an ALU-based hardware implementation and affects the hardware cost significantly. Figure 5-14 shows the classification accuracy against the number of bits used for FEE and the fully connected DNN classification unit (FCDNN-CLU). The classification results were initially calculated using full precision (32 bits) FEE and CLU. The number of bits for FCDNN not only affects the hardware complexity of FCDNN-CLU but also significantly impacts the memory requirement for the classification parameters (weights and biases). The FEE requires single-precision floating-point numbers and if the number of bits is reduced, the classification accuracy

degrades significantly as highlighted in Figure 5-14 (a). However, the number of bits for the DNN classification unit was reduced to half-precision (16 bits) without any significant effect on the classification accuracy as shown in Figure 5-14 (b). Therefore, we used a 32-bit FEE and a 16-bit classification unit for our NEOB prediction processor (DBE^2) to reduce the memory requirements by $\approx 50\%$. The implemented DNN architecture reduces the area (gate count) for the processor from 69.3K to 19.8K. Moreover, it achieves an overall energy efficiency of 1.13 TOPS/W for 16 bits precision, which is comparable to [96] with the integrated AFE.
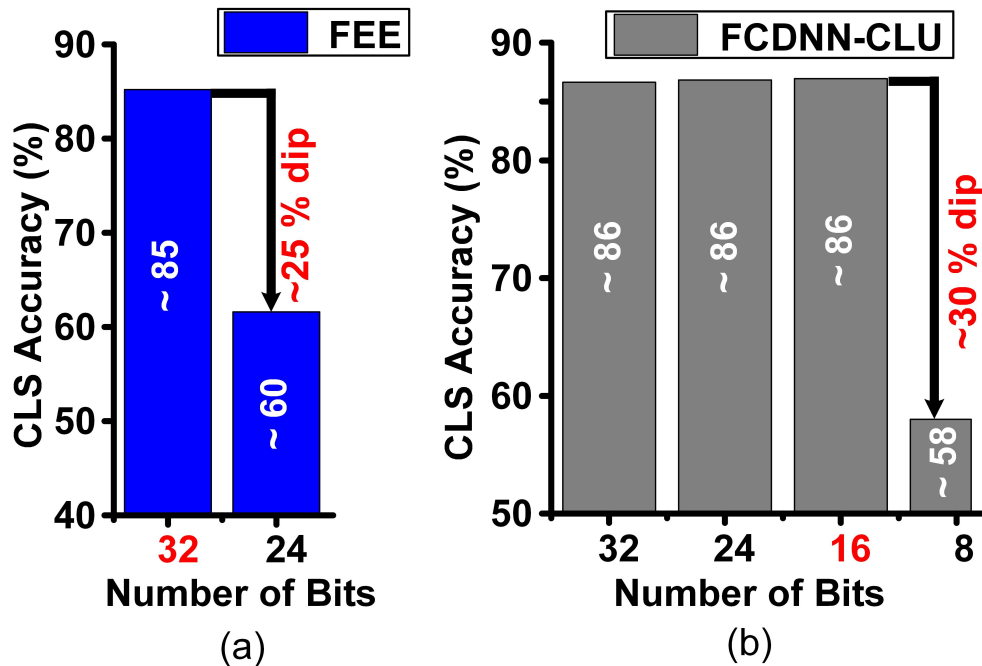
Figure 5-14: Classification Accuracy vs. Number of Bits for (a) Feature Extraction & (b) Classifier.

5.3 Measurement Results and Performance

The chip micrograph and the performance summary of the implemented DBE processor (DBE^2) for the NOEB prediction system are shown in Figure 5-15. The implementation fully integrates a 2-channel TCSCS AFE, SAR ADC, and 2-channel DNN processor for NEOB prediction. The SoC consumers an overall area and energy of $16mm^2$ and 10.13 μ J/classification, respectively, while implemented using 0.18 μm One-Poly-Six-Metal lay-

ers CMOS process. The SoC continuously tracks the NEOB and can provide real-time assistance to the caregiver/neurologist in assessing the emotional state in response to any learning activity for autistic people and beyond.

The total power consumption of the DBE^2 synthesized on a 1 MHz clock frequency is 1.29 mW, which down-scales to $\approx 1320\mu W$ with a 1 kHz clock frequency. The power of a system is defined as the energy per unit of time. Therefore, the energy utilization of the DBE^2 per prediction can be calculated by multiplying the power by the number of time units consumed per prediction, which is approximately $10.13\mu J/prediction.$

The estimated backup time of a battery can be calculated as equation 5.16. The estimated battery time in hours for DBE^2 processor using a coin cell battery of size, voltage, and area of 20 mm, 3 V, and 250 mAH, respectively is calculated in equation 5.17, which is ≈ 580 Hrs. Considering the aging factor and other losses of 50%, the estimated time would be ≈ 288 Hrs or 12 days.

$$Estimated\ Backup\ Time(Hrs) = \left[\frac{Battery\ Capacity\ X\ Input\ Voltage}{Total\ Load}\right] \quad (5.16)$$

$$Estimated\ Backup\ Time(Hrs) = \left[\frac{250mAH\ Capacity\ X\ 3V}{1320\mu W}\right] \quad (5.17)$$

Implemented Results	
Process	TSMC 0.18um 1P6M CMOS
Area	4.0 x 4.0 mm
Supply Voltage	1.0V (AFE) 1.0V (DBE)
NEF/channel	2.71
Classifier	DNN
Energy Efficiency	10.13uJ/Class.
Accuracy	85.2%
No. of Classes	4
PS Weight Memory	16KB
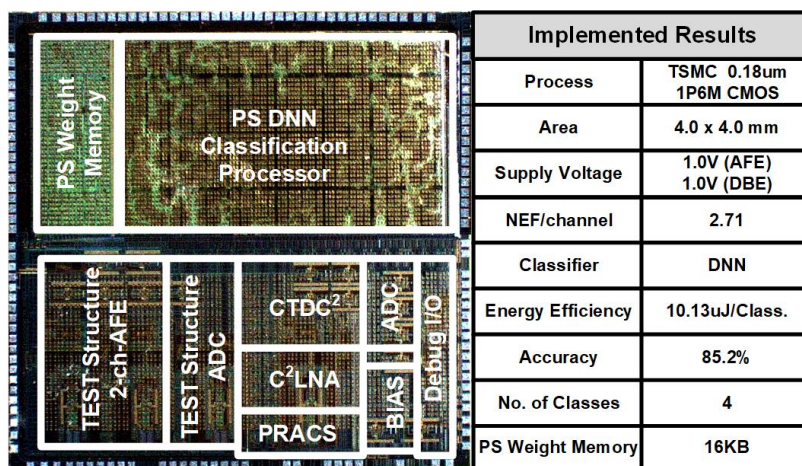
Figure 5-15: Chip Micrograph & Performance Metrics of the System.
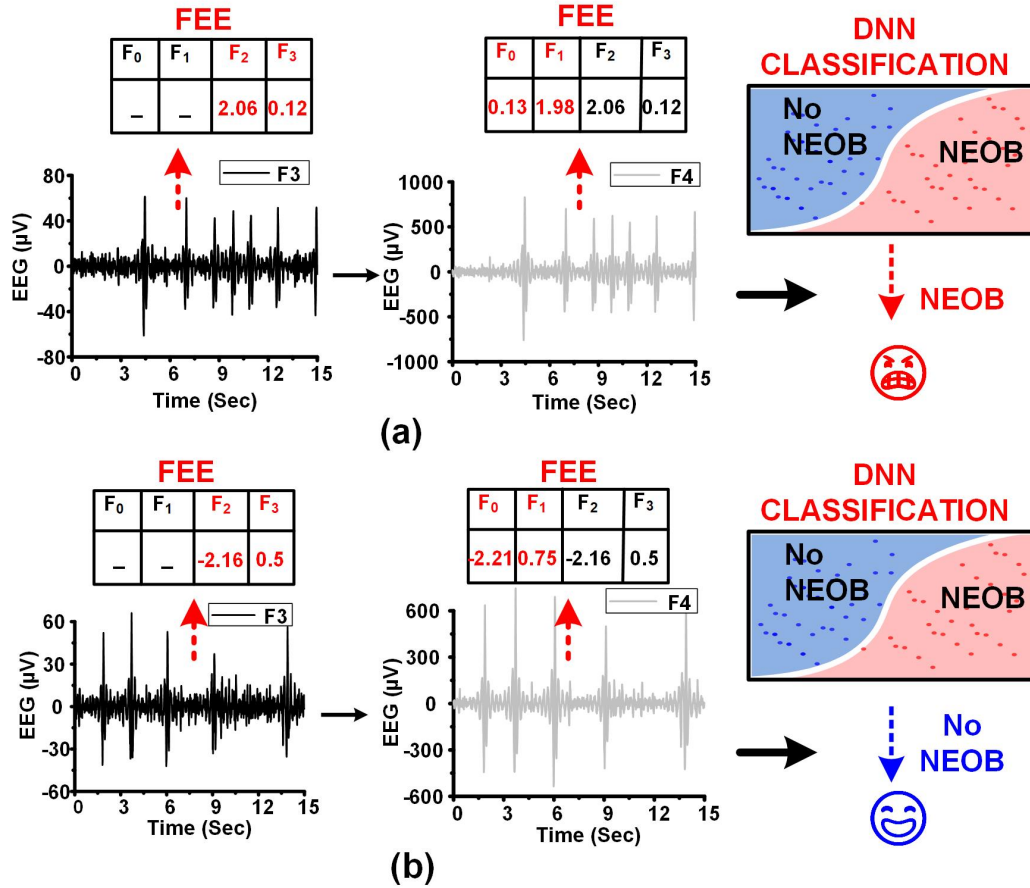
Figure 5-16: Measurement Results for (a) Negative Emotion Outburst & (b) No Negative Emotion Outburst.

Figure 5-16 shows the measurement results of the DBE processor for a subject (Subject #2, DEAP) for an emotion corresponding to NEOB and without NEOB. Both the multiplexed channels are processed in the same manner by the NEOB classification processor. The DNN classification unit classifies the EEG signal to predict a NEOB using the feature vector. The NEOB and No NEOB relate to emotion numbers 32 and 11, in the data set corresponding to [valence, arousal] values of [1, 9] and [7.08,1], respectively.

The hardware costly features were implemented in a customized manner to optimize the gate count to 86X lesser than conventional realization. The DNN classification unit was also optimized in a semi-pipelined manner and rescheduling the pipelined instructions using a customized ALU with 34X lesser resources. The SGM activation function was implemented using customized LUT, which provided an overall reduced energy consumption of 77% for the DNN classification unit.

Figure 5-17 shows the classification results for the NEOB prediction using the proposed NEOB prediction processor (DBE^2) for DEAP and SEED data sets. Figure 5-17 (a) shows the classification results of 16 subjects in the DEAP data set for NEOB prediction. Figure 5-17 (b) shows the classification results for valence classification of 15 subjects in the SEED data set. The classification results were evaluated using a leave-one-out cross-validation scheme. The overall sensitivity and specificity for NEOB classification were 78.80% and 89.10%, respectively.
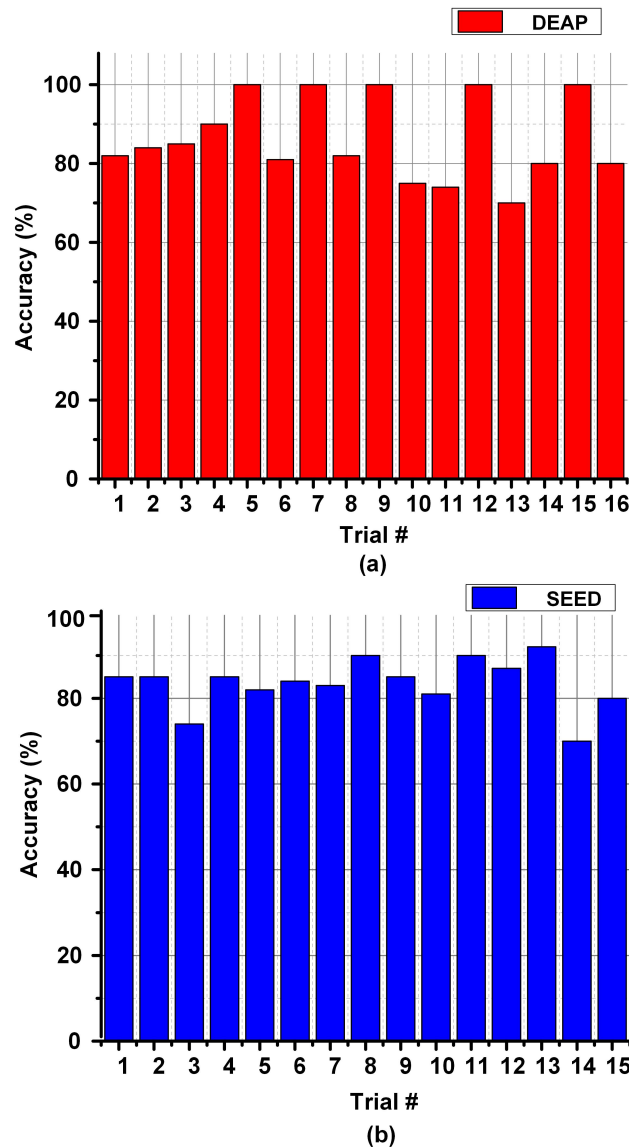
Figure 5-17: Classification Results for (a) DEAP (b) SEED.

Figure 5-18 shows the SoC measurement of an example for the NEOB test with an

analogous approach to the DEAP data set. The DNN was trained using > 100 different trials of emotions with NEOB and without NEOB from the emotion classification data sets. Short videos (1–2 min) with happy (No NEOB) and sad (NEOB) contents were displayed to the subject (Age: 7 Years, Gender: Male) with learning disability signs. Two-channel EEG information was recorded simultaneously, the features were extracted, and DNN was trained offline. The top part of the figure depicts an example to show the EEG signals for a trial of happy emotion followed by two trials of sad emotion.
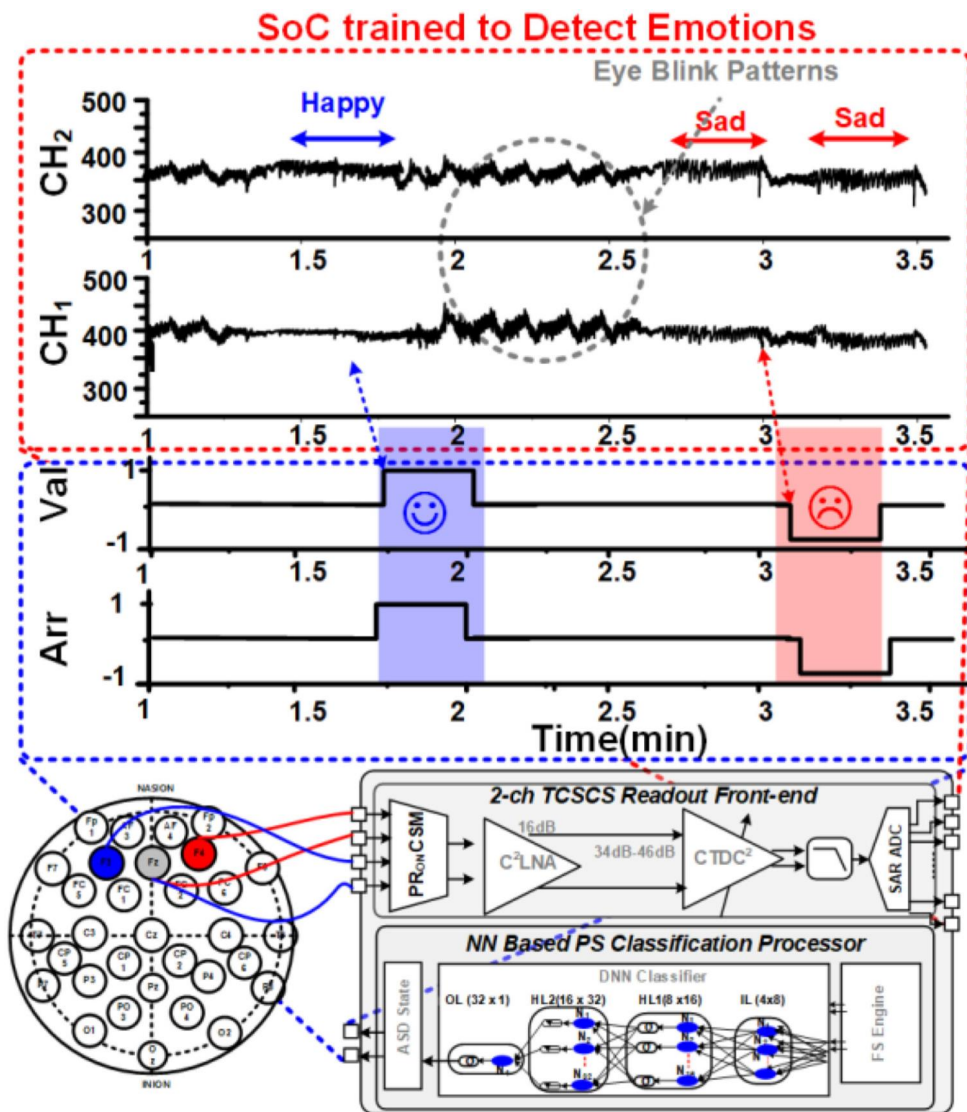
Figure 5-18: SoC Measurement for Negative Emotion Outburst Prediction.

To verify the full functionality of the SoC, we uploaded the extracted DNN parameters

on SoC and, repeated similar video content to evaluate the real-time performance. The NEOB DNN classifier was successful in classifying the majority of them and achieved an average accuracy of $> 90\%$ for more than 15 min of continuous operation of the SoC.

5.4 Performance Comparison

The proposed DBE^2 DBE processor was the second fully integrated on-chip processor for NEOB prediction, after our 1st negative emotions prediction processor DBE^1. The performance comparison of the DBE^2 NEOB prediction processor with the previous state-of-the-art works are summarized in Table 5.1.

H. Gonzalez [97] performed the emotions classification using 14 EEG channels. They achieved the classification results of 72.4% using a convolutional neural network (CNN) classifier on two public BM-DS including DEAP and DREAMER. They also did not provide any hardware implementation, and only provided a software-based system. The number (14) of EEG channels utilized by them was also high, and hence not feasible for a wearable SoC implementation.

The first negative emotions prediction processor (DBE^1) of this Ph.D. project predicted the negative emotions with 73.4% classification accuracy [38]. The DEAP and SEED public BM-DS were utilized for the system validation. It was the first fully integrated SoC solution, which utilizes 8 EEG channels for negative emotions prediction. Although the classification results of this system were very good, but not high ($< 80\%$). The number of EEG channels and the classification energy ($16\mu J$) were also high. This SoC was designed using 180 nm TSMC CMOS process.

H. Gonzalez [25] proposed an emotions prediction processor titled "BioCNN" using a CNN. The system utilizes 14 EEG channels for emotion prediction with 83.12% classification accuracy. The processor was implemented on a Digilent Atlys Board with a Spartan-6 FPGA. An excellent energy efficiency of 11 Gops/W was provided by the system. However, the system lacks a fully integrated SoC, and only an FPGA implementation was provided. The number of EEG channels was also high and not suitable for a wearable SoC implementation.

T. Song [45] proposed an emotions prediction algorithm using an A-LSTM recurrent neural network. An emotions prediction data set MPED was used in this work. The

Table 5.1: Comparison With the State-of-The-Art Works.

	H.Gonzalez [97]	A.Aslam [38]	H.Gonzalez [25]	T.Song [45]	S.Koelstra [24]	This Work [39]
Hardware	No	Yes	Yes	No	No	**Yes**($0.18\mu m$)
Area	—	—	$3.35mm^2$	—	—	**16** mm^2
Power/Energy	—	—	76.61mW	—	—	$10.13\mu J$
Data Set(s)	DEAP	DEAP	DEAP	DEAP,SEED	MAHNOOB	**DEAP, SEED**
Ch. Count	32	Variable	6	62	32	**2**
Classifier	NB	SVM	CNN	GELM	LSTM	**DNN**
Accuracy (%)	59.8	73.8	80.1	80.4	82.5	**85.2**
Multiple BM-DS	No	No	No	Yes	No	**Yes**

provided an emotions prediction algorithm with 81.1% accuracy. Although they provided excellent classification results, but the number of EEG channels utilized by them is very high. They also did not provide any hardware implementation of the system.

W. Fang [26] proposed an SoC-based emotions prediction classifier using 28 nm CMOS process. They implemented a CNN-based classifier for emotions prediction using 6 EEG channels with 80.1% accuracy. The area and power consumption of their SoC were 3.35 mm^2 and 76.61 mW respectively. The algorithm was validated on the DEAP data set. The main limitations of their system were the lack of a fully integrated system. The features were extracted offline and then forwarded to the SoC for emotion prediction.

The previous systems are either software-based, FPGA-based or lack a fully integrated SoC solution. The proposed DBE^2 DBE processor for emotion prediction provides a fully integrated SoC solution using a minimum number of EEG channels. The system provided excellent classification results (85.2%) with a low area ($16mm^2$), and minimum energy consumption ($16\mu J$). The SoC is validated using multiple BM-DS including DEAP and SEED data sets.

5.5 Summary

In this chapter, I have explained the design of DBE^2 negative emotions prediction processor. The DBE^2 negative emotions prediction processor has the following novel points:-

- Second Fully integrated twin-channel SoC based NEOB prediction processor.

- Utilizes a fully connected DNN classifier for NEOB prediction.

- Continuous NEOB prediction can help parents or caregivers to suppress this NOEB.

- Processor's performance is validated using well-known BM-DS inluding DEAP and SEED.

- Utilizes an an area-and-energy-efficient feature extraction including ZC and ASKI.

- The proposed approximated FV implementation reduced the overall area (gate count) by 86X.

- An area-and-energy-efficient semi pipe lined DNN architecture was used for DNN implementation.

- The semi pipe lined DNN consumed 77% lower classification energy.

- A LUT based SGM unit was implemented for DNN with 50% lesser memory resources.

- Fabricated using TSMC 180nm CMOS process.

- Provide excellent ($> 85\%$) classification results.

- Utilizes minimum (2) number of EEG channels among hardware based EEG classification systems.

5.6 Conclusion

The proposed DBE^2 processor consumed an active area of $16mm^2$ while consuming an energy of $10.13\mu J/classification$. The proposed processor can be embedded inside a hand band SoC after integration with ULP and low-noise AFE. The classification processor would be operable with a coin cell battery of area, voltage, and current of 20 mm, 3 V, and 250 mAH respectively, for a time duration of one week to provide a real-time

emotional prediction. The proposed DBE^2 negative emotions prediction processor provided excellent (85.4%) classification results for NEOB prediction. The classification energy ($10.13\mu J/prediction$) and the number of EEG channels (2) for the NEOB prediction were also minimum in comparison with the state-of-the-art works.

I have achieved excellent classification results with the lowest number of EEG channels for real-time emotion prediction. During the design and development of the DBE (DBE^1 and DBE^2) processors for real-time negative emotions prediction, I noted that the selection of a minimum number of suitable channels, and features are a major challenge for the researchers for emotions prediction using EEG signals. The choice of channels and features in previous research is either based on the previous literature or some limited self-experimentation. I have, therefore, performed a LSFE to identify the best suitable channels and features for emotion prediction. This suitable channel and feature identification would be highly beneficial for the DBE processors for emotion prediction. I have also performed a LSFE analysis for ASD prediction using EEG signals. The next chapter (Chapter 6) explains the LSFE methodology for emotions and ASD prediction using EEG signals.

Chapter 6

Suitable Channels & Features Identification: A LSFE Methodology

The recorded EEG signals are contaminated by noise and other artifacts including eye movements, eye blinks, muscle activities, and chewing [98]. The complete set of EEG signals also contains a very large amount of data due to the large number ($\geq$16) of channels. The presence of noise and the size of data makes it difficult for the ML classification algorithms to accurately map the input (EEG signals) to the output (label). Therefore, it is desired to remove the noise and other artifacts to identify the most relevant subset of the information required for the classification.

Channel selection is the process to identify the most suitable subset of channels out of the total number of available channels. The feature extraction process is to identify some variables or formulas instead of raw EEG signals for classification. The channel selection process selects a subset of the most suitable channels. The feature extraction process calculates features from these channels and the FS process selects the best subset out of these features.

The channel and FS are iterative processes and one of the most challenging procedures in ML classification. There is no fixed method to identify the best suitable features and channels for a time series EEG classification problem. Different features have to be exper-

imented by the researchers to find the best suitable features and corresponding channels. The channel and FS are not only required for the accurate prediction or classification of the label. But they also directly affect the hardware resources for hardware accelerators or on-chip systems [89]. Generally, there are two approaches used for the channel and FS process a) self-experimentation using previous literature and domain knowledge, and b) LSFE. The LSFE techniques calculate a huge number of features from the input data, which are, later on, analyzed to filter for the best suitable features.

The self-experimentation approach is usually considerably limited and analyzes a significantly smaller number of features as compared to the LSFE method. To the best of our knowledge, there is no previous work that has utilized the LSFE to identify the most optimal channels and features for emotions and ASD classification. Therefore, LSFE to identify the most suitable channels and features for emotions and ASD classification using DEAP, SEED, Old Dominion University (ODU), and King Abdul Aziz University (KAU) data set was utilized.

The DEAP and SEED data sets for emotion prediction are already discussed and explained in Chapter 2 (Section 2.4). The DREAMER data set provides the EEG and ECG data of 23 participants. The EEG signals are recorded using 14 electrodes through an emotive EPOC headset [99]. It provides the labels of valence, arousal, and dominance for the emotion's classification. We have utilized the labels of valence and arousal only from this data set. The DREAMER data set is unevenly balanced between both genders. The valence and arousal labels in the DREAMER data set are scaled between 1 and 5. The valence and arousal values lesser or equal to 3 are labeled as low, and those greater than 3 are labeled as high. It can be observed that the binary classes of arousal are evenly balanced using a threshold of 3. However, the valence classes are unevenly balanced in the data set.

A brief overview of the ASD prediction data sets is provided below before large-scale feature extraction methodology for emotions and ASD prediction.

6.1   ASD Classification Data Sets

There are very few ASD classification data sets compared to other problems like emotional classification, epilepsy, Alzheimer, etc. The primary reason for that can be the uncooperative behavior of ASD children, as they are generally very uncooperative. The uncooperative

nature of ASD children is due to multiple reasons, a) difficulties to understand instructions, b) social interaction and c) communication and sensory issues. The EEG classification data sets for ASD classification records the EEG data for a certain number of ADOS-2 confirmed ASD patients and TD children.

The ASDOS-2 is a standard ASD diagnosis method. It requires a communication score (CSC), social interaction score (SCI), imagination and creativity score (IMC), stereotyped behaviors score (STB), and their cut-off values with a certain threshold table being evaluated by the neurologists [100]. A behavioral diagnosis cycle of the patient evaluates the SCI, CSC, STB, and IMC scores of the patient. The scores are compared with the cut-off table and then the patient is labeled as ASD or TD [100].

The EEG-based ASD classification system is trained to predict the patient as ASD or TD using EEG signals. The predicted labels are then compared with the original labels assigned through the ADOS-2 method [100]. There is no publicly available EEG-based ASD data set to the best of my knowledge. I have utilized the ASD data sets by Old Dominion University (ODU), USA and KAU, KSA shared with me for this research [101]-[102].

ODU data set provides the EEG data of 17 subjects including 8 ASD and 9 TD subjects. The subjects include 10 males and 7 females [101]. The males include 6 ASD and 2 TD subjects. The females include 2 ASD and 5 TD subjects. The ODU data set has also provided the ADOS-2 scores of each patient on which they were labeled as ASD or TD. The ASD patients have a higher ADOS-2 score than the TD subjects. The ODU data set recorded the EEG signals of all subjects using 32 electrodes. However, in our analysis, we observed that the EEG signals for only 14 electrodes (F7, F3, Fz, F8, FC1, FC2, FC6, T9, T7, C3, T10, CP5, CP2, P7, P3) were available across all subjects. The EEG data of the remaining electrodes were either missing in some subjects or too noisy to be included in the analysis. Therefore, we have focused only on these 14 channels in this work.

KAU data set provides the EEG data of 12 children including 8 ASD children and 4 TD children [102]. The data set does not provide the ADOS-2 scores of the patients. The ASD children include 5 boys and 3 girls whereas the TD children include 4 boys. All the participants were aged between 10 to 11 years. They recorded the EEG data using 16 channels (FP1, FP2, F7, F3, Fz, F4, F8, T3, C4, Cz, C3, T5, Pz, O1, Oz, and O2).

6.2 LSFE AND CLASSIFICATION METHODOLOGY

The feature extraction process for the time series classification problems, including emotion recognition and ASD prediction, is hectic [103]. It requires the analysis of a large combination of features using previous domain knowledge and experimentation with different features. Some LSFE packages including TSFRESH, TSFEL, HCTSA, etc using MATLAB or python implementations are proposed by different researchers [104],[105].A large set of features using the time series EEG data is calculated by these packages. These features can be passed through different feature selection (FS) methods, including selecting k best (SKB), sequential forward search (SFS), etc. to select the best optimal feature subset. The FS methods utilize different learning algorithms to find the best subset of features from the large feature set acquired through the LSFE method.

The block diagram of the LSFE methodology for emotions or ASD prediction is depicted in Figure 6-1. The complete set of prepossessed EEG data is passed through the threshold process for label creation if required. If the data set has already provided the binary labels for positive and negative emotions or ASD/TD classification, then the threshold process is not required. The EEG data and labels are forwarded to the LSFE method (TSFRESH etc) for LSFE analysis. A large set of features is calculated by the LSFE process. The LSFE features are forwarded to the channel and feature selection methods. The small subset of features after suitable channel and feature selection is used for the preparation of the feature set. The final feature set after feature preparation is provided to the classification method (CLS) for positive/negative emotions or ASD/TD classification.

The TSFRESH package for the LSFE of emotions classification and ASD prediction was utilized in this work. The TSFRESH package provides 63-time series characterization methods and calculates a total of 794 features for each EEG channel. It can be used for both univariate and multivariate time series and can handle variable-length time series. Some of the features used in TSFRESH include mean absolute energy, absolute maximum, number of peaks, quantile value, and zero crossings. An exhausting list of features extracted in TSFRESH is listed in [106].

The large-scale feature matrix is then forwarded to the FS method to find the best subset of features.The SKB and SFS methods are used in this work [107]. SKB is an implementation of a filter FS method where it eliminates all excluding the highest-scoring
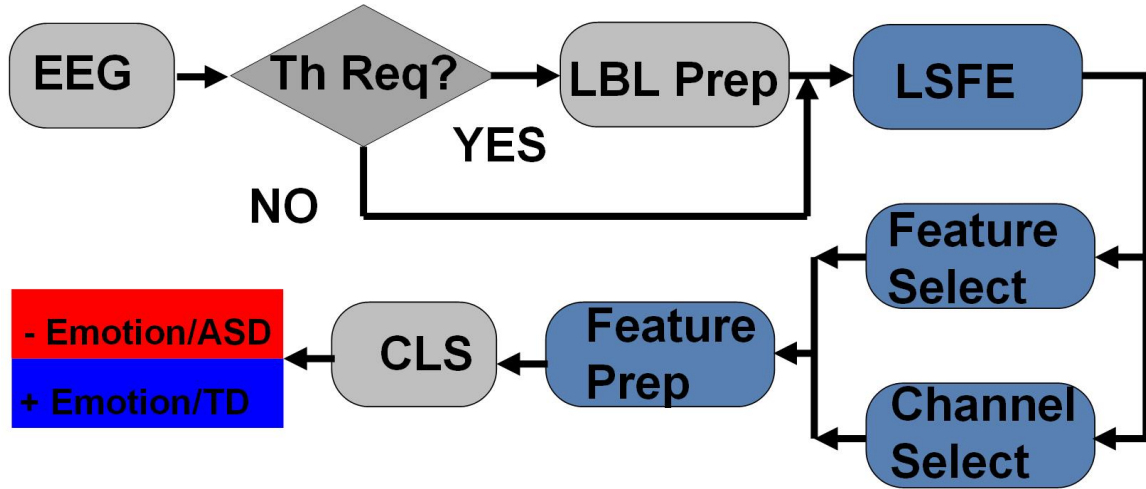
Figure 6-1: Large Scale Feature Extraction & Classification Methodology.

features. The number of selected features is controlled by the k value. The SKB uses a linear method to select features that contain information about the target variable using statistical tests such as ANOVA, Fisher score, and Chi-Squared [108]. The selection of important features was based on the ANOVA test with k=8. SFS is an iterative wrapper-based method [108]. It starts with an empty set and adds features to form a feature subset. Those selected features give the highest value for the objective function. The objective function is defined by a perceptron in our case. The FS is a back-and-forth learning process through the relevant learning algorithm. The FS methods work to reduce the size of the feature matrix. They reduce overfitting by excluding redundant features, improving the classification results, and decreasing the classifier's training time. LSFE and FS is performed for the selected data sets. The results of valences in emotions classification data sets are presented in detail, whereas the results of arousal classification are briefly summarized.

6.2.1 Emotions Classification LSFE

The DEAP and SEED data sets include the data of 32 and 62 EEG channels, respectively. Several previous works have identified different smaller pools of the most suitable channels for emotion classification with customized feature sets [24]. However, a detailed analysis of all EEG channels were not performed to show the significance of their channel subsets. The utilization of different channels using LSFE to identify the best suitable channel subset was analyzed in this chapter.

The LSFE matrix calculated a considerable number ($\approx$25K) of features for the DEAP data set using 32 EEG channels. The classification performance using the LSFE matrix was not satisfactory ($\leq$65%). The primary reason for the low classification results using LSFE matrix with all EEG channels was over-fitting and redundant features. The SKB and SFS methods significantly improved the classification results (85%) for emotions classification. Figure 6-2 represents the classification performance of LSVM, SNN, KNN, DT, and XGB classifiers using the LSFE matrix before and after FS. Figure 6-2 (a) represents the classification results using the LSFE matrix. Figure 6-2 (b) represents a box chart of the subject-wise classification results after FS for all subjects in the DEAP data set. The red and blue colors in Figure 6-2 (b) indicate the classification results of SKB and SFS methods, respectively. It can be observed that the classification results were significantly improved.
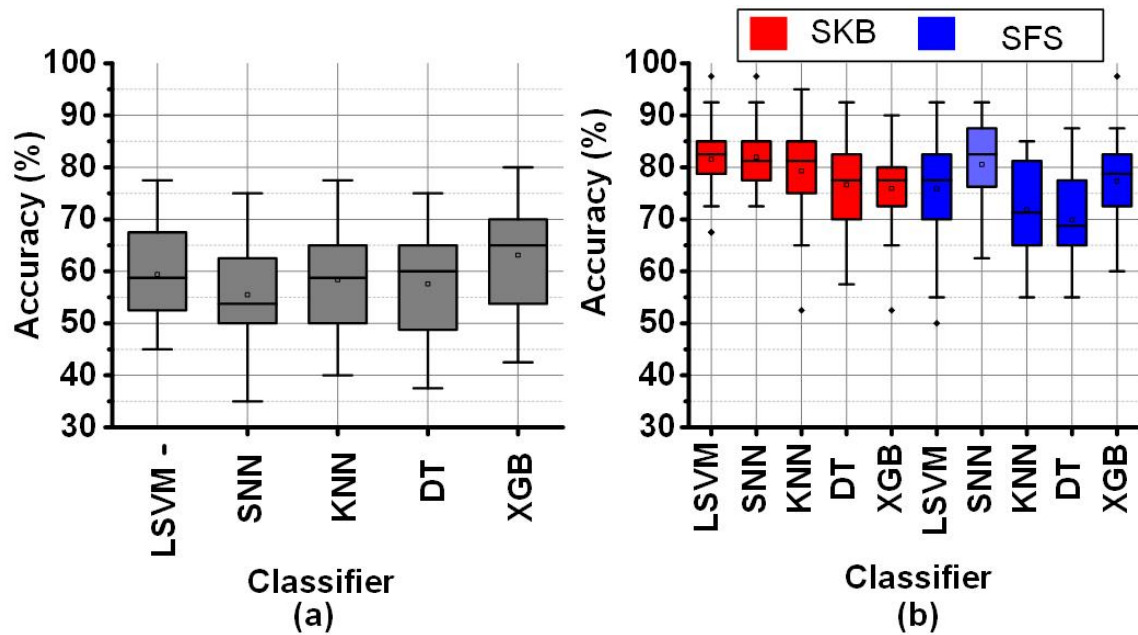
Figure 6-2: (a) Large Scale Feature Extraction Classification (b) Feature Selection Using Select K Best & Sequential Forward Search for DEAP.

The identification of a smaller subset of suitable channels was performed using the SKB and SFS methods. These FS methods were tuned to utilize a set of 8 best-suitable features for the classification. As a result, the FS methods improve the classification performance and significantly impact (>3k times) the hardware resources for hardware accelerators and on-chip applications. Figure 6-3 represents the utilization of each EEG channel using the

SKB and SFS methods. The heat maps in Figure 6-3 (a) and Figure 6-3 (b) represent the channel importance using SKB and SFS, respectively. It can be observed in Figure 6-3 (a) that channel number 1 (FP1) is utilized for all features (8/8) using the SKB method for subject number 11.
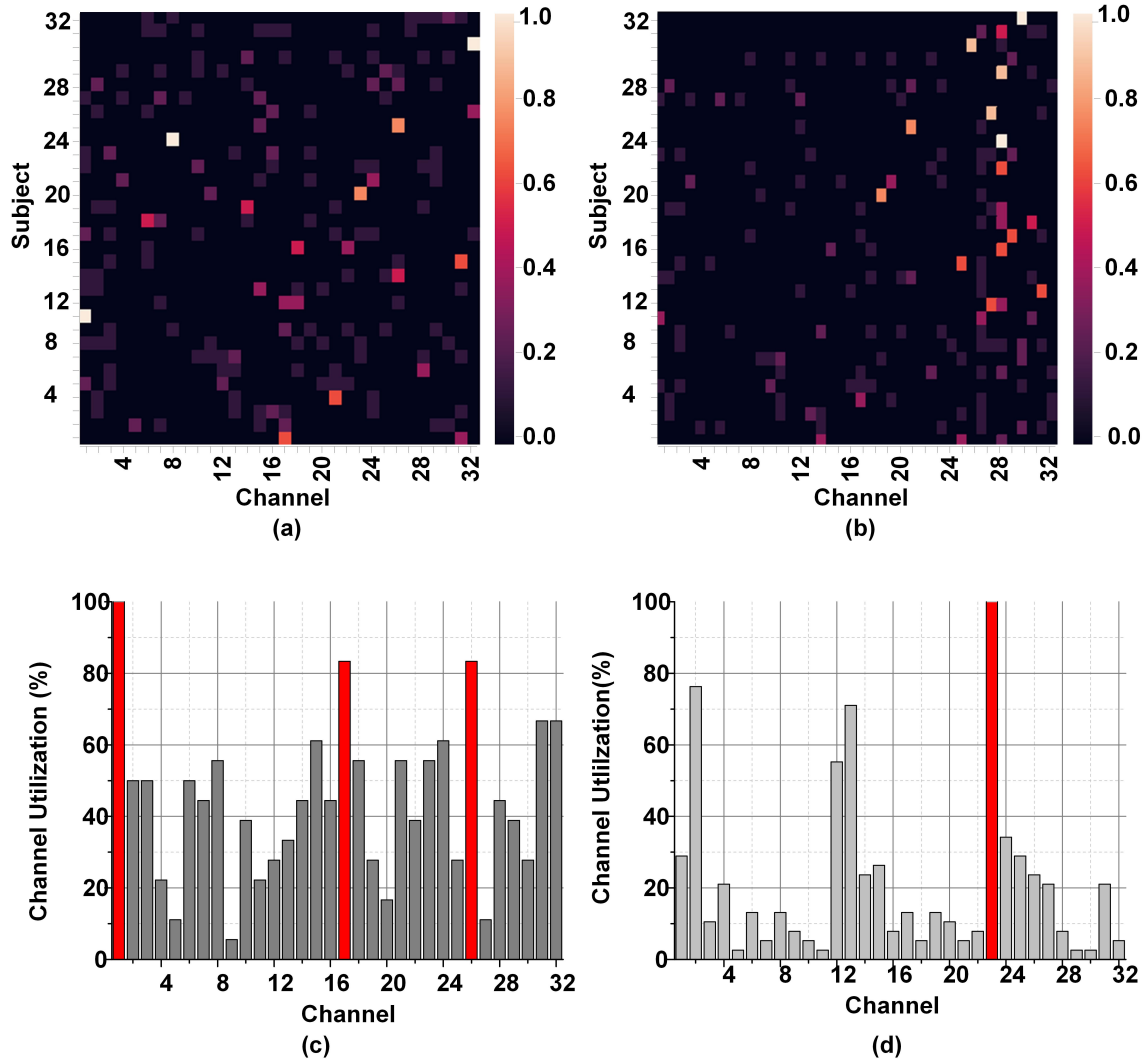
Figure 6-3: Heat map for Feature Selection using (a) Select K Best, (b) Sequential Forward Search, (c) Bar Graph for Feature Selection using Select K Best, & (d) Sequential Forward Search for DEAP.

The maximum (8/8) and minimum (0/8) utilization of the channel are reflected by white and black colors in the heat map respectively. However, it was difficult to completely analyze the utilization of each channel for each subject using the heat maps. Therefore, we have presented the accumulated utilization percentage of each channel using SKB and SFS

in Figure 6-3 (c) and Figure 6-3 (d), respectively. These bar graphs show the accumulated utilization of each channel for all subjects in the DEAP data set. It can be observed that channel numbers 1 (FP1),17 (FP2), 23 (FC2), and 26 (T8) were most suitable for the emotion classification. The T8 channel was also observed to be among the eight-channel pool in our previous work [72]. This channel subset relates to the amygdala region of the brain [109]. The amygdala region of the brain is closely linked to human emotions [110]. After LSFE, the most significant features identified using these FS methods include autocorrelation, Fourier transform coefficients, signal energy, continuous wavelet transform coefficients, change quantiles, and aggregated least square regression.

The LSFE feature extraction and FS methods improved the classification results for the SEED data set significantly compared to the DEAP data set. Figure 6-4 presents the classification results of SEED after FS and LSFE using TSFRESH. The red and blue colors represent the SKB and SFS methods, respectively. Both FS techniques provided excellent classification results (98%) for SEED data set.
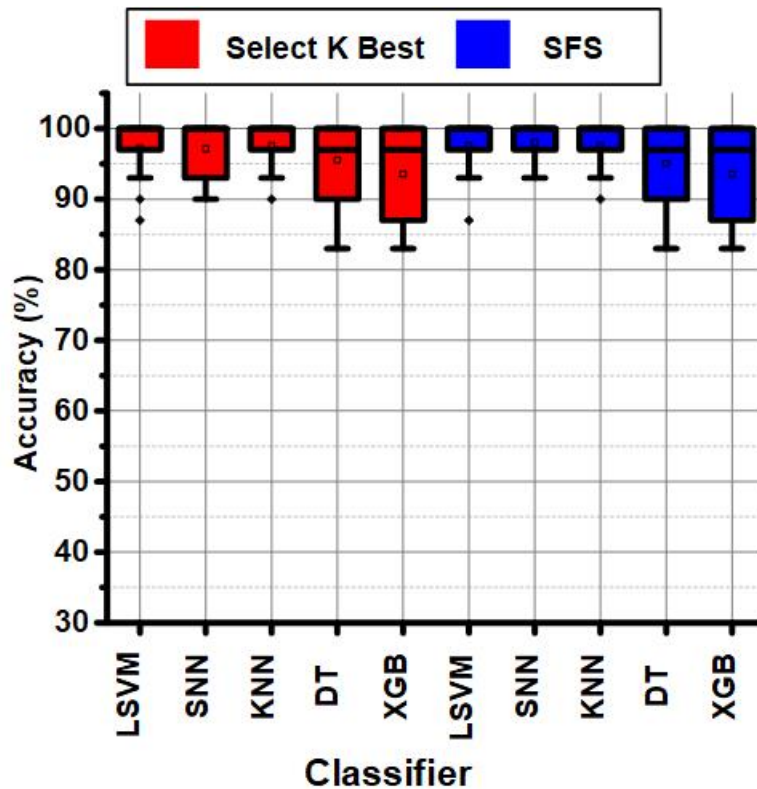
Figure 6-4: Classification Results After Feature Selection for SEED.

The identification of a smaller subset of suitable channels (4 channels) for the SEED and DREAMER data sets would reduce the hardware resources to ≈12.5k and ≈3.5k times, respectively. Figure 6-5 shows accumulated channel utilization for SEED using SKB and SFS methods. Figure 6-5 (a) and Figure 6-5 (b) represent the channel importance for the SEED data set using SKB and SFS, respectively.
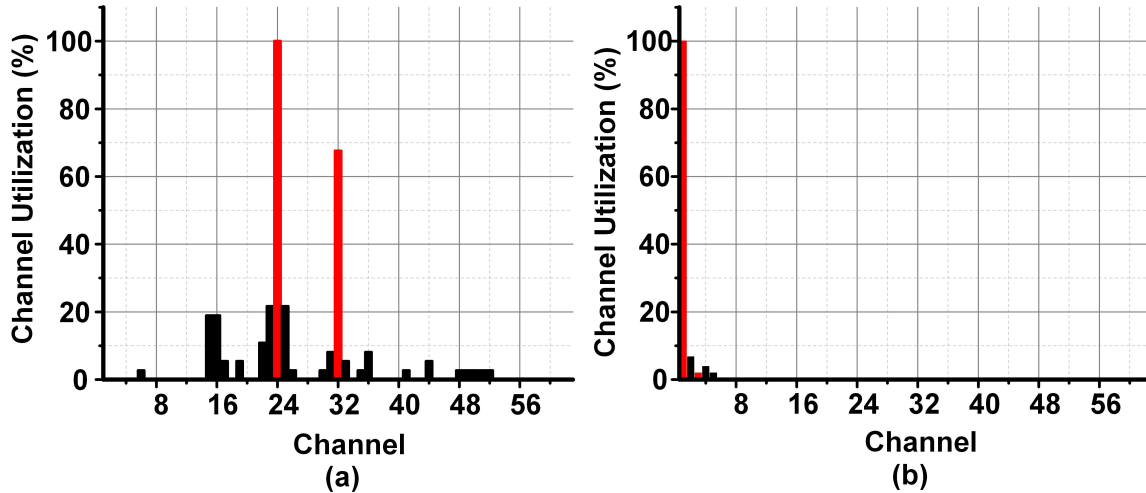
Figure 6-5: Channel Importance for SEED using (a) Select K Best, & (b) Sequential Forward Search.

The channel utilization analysis suggested that channel numbers 24 (T7), 32 (T8), and 1 (FP1) were significantly important in the SEED data set. It can be observed from Figure 6-5 (a) and Figure 6-5 (b) that only three EEG channels have significantly important utilization. The channel combination (T7, T8) confirmed our previous observations of the significance of temporal channels and asymmetric electrodes combination for emotions classification [89]. Channel number 3 (FP2) was included as the fourth channel as a 4-channel subset was targeted. This channel selection was used to exploit the benefit of asymmetric electrodes combination [89]. The channel combination (T7, T8, FP1, FP2) further improved the classification results as discussed later in this section.

6.2.2 ASD Classification LSFE

The ODU ASD classification data set provides the EEG signals of 14 channels for all subjects as explained earlier in Section 6.1 [101]. The LSFE matrix provided ≈11k features for the ODU data set. The LSFE matrix provided a maximum accuracy of 83% for the ASD

classification (KNN classifier).

Figure 6-6 (a) and Figure 6-6 (b) show the classification results for ASD classification for the ODU data set using LSVM, SNN, KNN, DT, and XGB classifiers before and after FS respectively. The SKB and SFS methods were used to extract the best 8 features from the LSFE matrix. The SKB methods provided excellent classification results (100%) whereas the SFS method provided the maximum classification accuracy of 93%. The FS methods would reduce the hardware complexity $\geq$ by 1.4k times for the hardware accelerators and on-chip applications [89],[72].
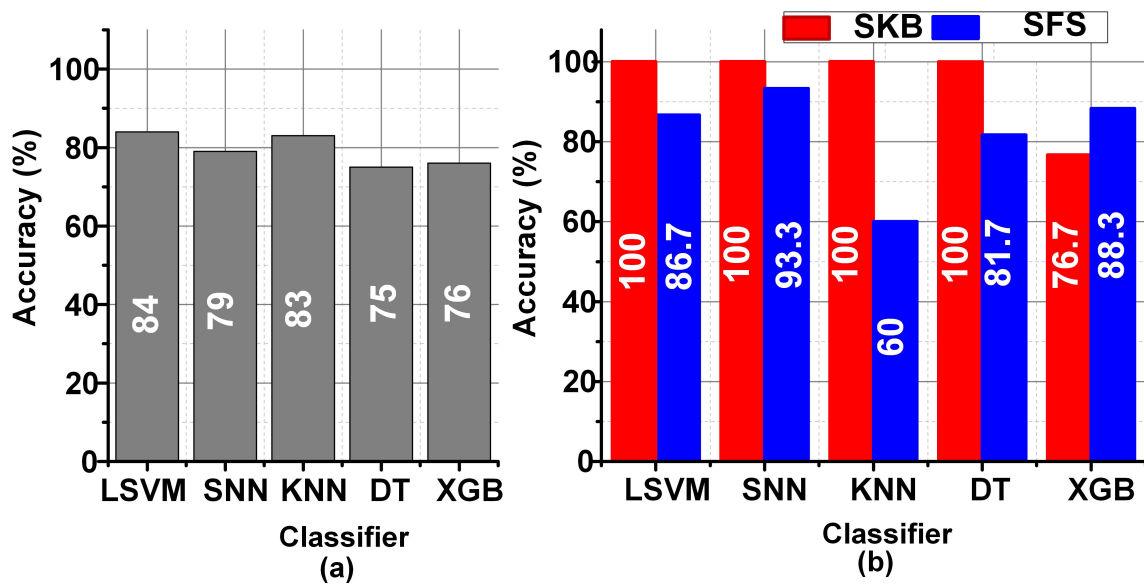
Figure 6-6: Classification Results for ASD (a) Before Feature Selection (b) After Feature Selection using ODU Data Set.

The identification of a smaller subset of suitable channels (4 channels) for ASD classification would also reduce the discomfort for ASD children for wearable on-chip applications. Figure 6-7 shows the channel importance of different channels for ASD classification with SKB and SFS methods using the ODU data set. The subset of 8 channels using SKB and SFS are shown in Figure 6-7 (a) and Figure 6-7 (b), respectively. It can be observed that the channel locations CP2, FC2, F7, and FC6 have the maximum utilization among the 8 feature subsets. Therefore, these channels were identified to be most suitable for ASD classification using the ODU data set.

The KAU data set provides the EEG signals of 16 channels for 12 subjects as explained
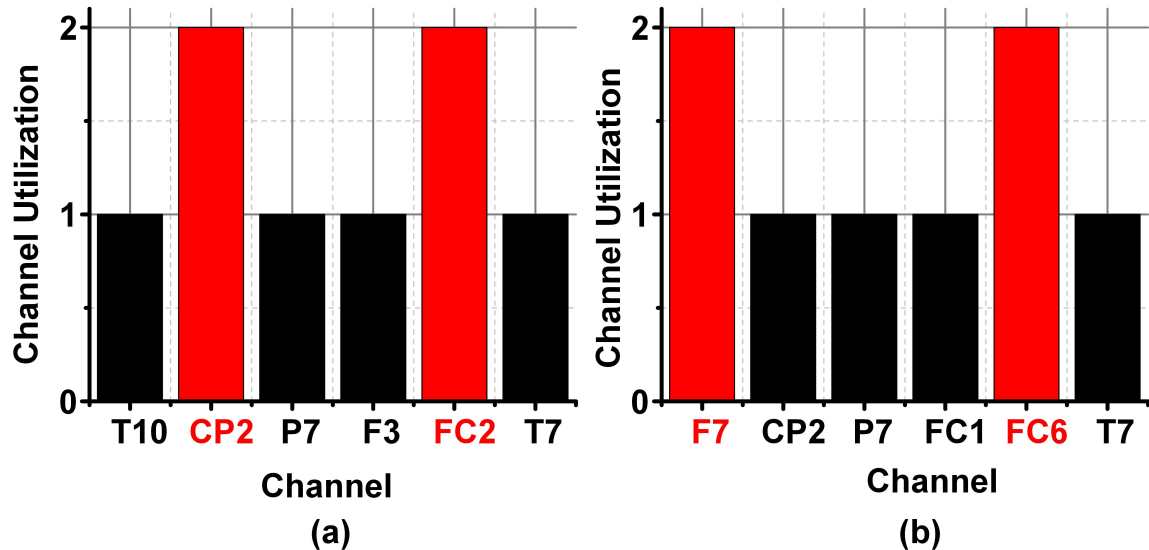
Figure 6-7: ODU Data Set Channel Importance for ASD Classification
After (a) Select K Best (b) Sequential Forward Search.

earlier in the data set section. The LSFE matrix provided ≈12.5k features for the KAU data set. The LSFE matrix provided a maximum classification accuracy of 77% for the ASD classification using the SNN classifier. Both the FS methods provided excellent classification results (≈ 100%) after FS. The FS methods would also significantly impact the hardware complexity ≥1.6k times for the hardware accelerators and on-chip applications.

Similarly, the channel locations F3, T5, O1, and O2 have the maximum utilization among the 8 feature subsets in the KAU data set. Therefore, these channels were identified to be most suitable for ASD classification using the KAU data set. The most suitable features for ASD classification include Fourier transform coefficients, auto-correlation, and mean absolute energy.

6.2.3   Channel and Feature Subset Settlement

The settlement or consensus of the identified channel and feature subset is required to conclude the suitable channel and feature subset for emotions classification. The primary reason for this is the subject-wise classification and dependence of the channel and feature subsets. The channel and feature combination proposed in this work provided the best classification results. However, due to the subject-wise dependence on emotions classification, the selected subject of channels may not correspond to the FS of many subjects.

117

Figure 6-8 (a) shows the channel utilization percentage of the best EEG channel (FP1) in the DEAP data set using SKB. The best feature subset using SKB does not contain any feature corresponding to the FP1 channel for multiple subjects. Therefore, it is impractical to have a subject-specific channel selection. But the classification results were degraded ($\approx 10\%$) using the selected subset of 4 channels.
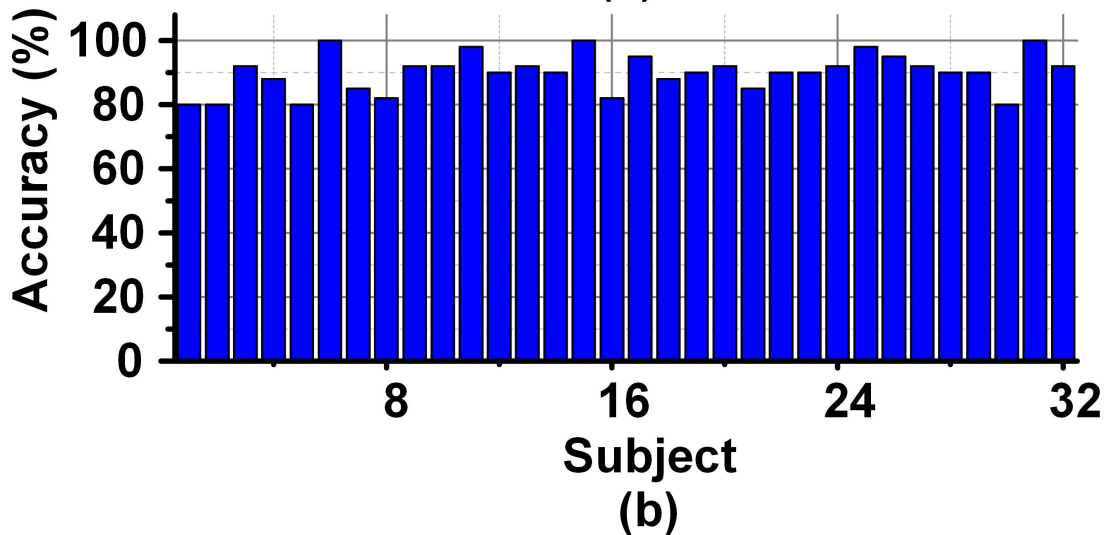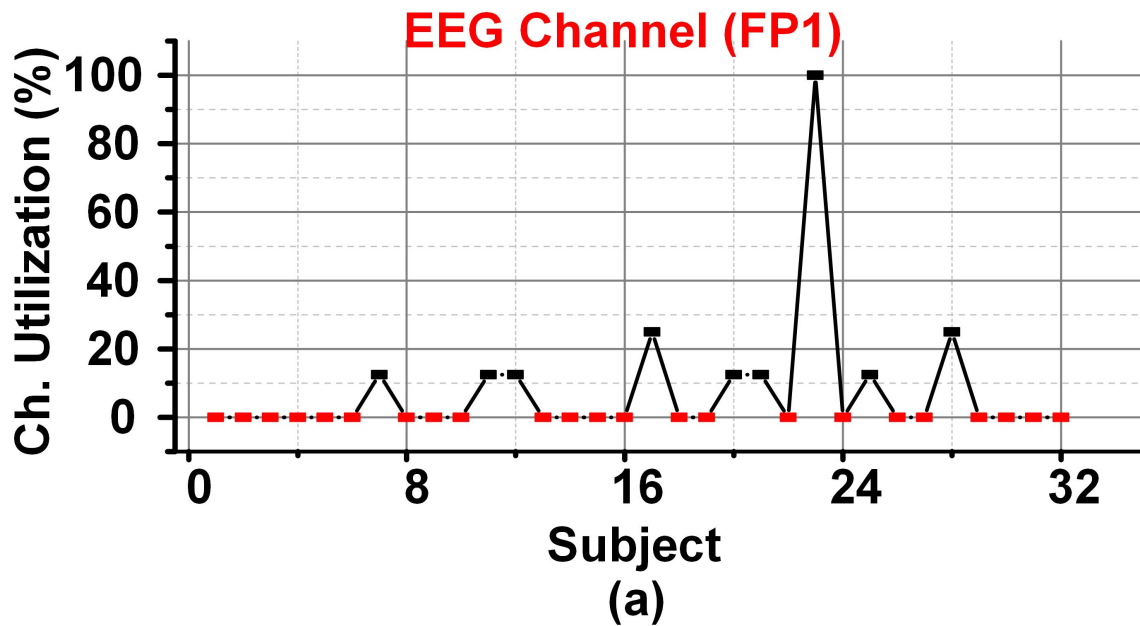
Figure 6-8: (a) Channel Importance of FP1 Channel in DEAP (b) Classification Results using Brute Force Approach in DEAP.

The primary challenge of the research for emotions classification using EEG signals is

to discover a subject-independent channel and feature subset feasible for the classification. Therefore, the LSFE and FS process to select the best 32 features using the subsets of 4 EEG channels using DEAP, SEED, and DREAMER data sets was repeated. The selected features include auto-correlation, change quantiles, entropy, signal energy, aggregated linear trend, welch density, Fourier entropy, and auto-correlation. A brute force approach was applied to select the best possible subset of 8 features from all possible ($\approx$10 million) features. The classification results from the brute force approach using the DEAP data set are shown in Figure 6-8 (b). The average classification results of 90.1%, 95.22%, and 98.56% were achieved using DEAP, SEED, and DREAMER data sets for valence classification using the LSVM classifier.

The LSVM classifier was selected based on the classification results and the classifier's complexity. A subset of eight features is forwarded to the classifier for positive and negative emotion or ASD classification. The dot product of the FV is calculated with the classifier weights as mentioned in equation 6.1. A positive or negative class is assigned based on the comparison of the dot product with a constant K. The LSVM process assigns the positive or negative class assignment to a feature based on the dot product.

$$C = FV.W \geq K \tag{6.1}$$

The complexity of the LSVM classifier is O(n), where n is the number of input dimensions [111]. The number of input dimensions is dependent on the number of features. The LSVM classifier is chosen due to the classification results and lower complexity than DNN or CNN. The DNN or CNN with the same or more input dimensions than an LSVM would always have a higher complexity than the LSVM due to activation functions. For example a shallow neural network with eight input nodes and one output node with a sigmoid activation function would have more computational complexity than an LSVM with 8 features. We have therefore preferred the LSVM classifier.

6.3   Results

Emotions and ASD classification and its analysis using DEAP, SEED, DREMAER, ODU, and KAU data sets were performed in this study. A total of 99 subjects were classified

and analyzed. The emotions classification included positive or negative valence and arousal classification for the DEAP data set. The SEED and DREAMER data sets were utilized for the positive or negative valence classification only. The DEAP data set provided the valence and arousal values from 1 (minimum) to 9 (maximum). A classification threshold of 5 was used in the data set to mark the labels as positive or negative. The SEED data set does not require a classification threshold as the positive or negative valence labels were provided by the data set. The valence values from 1 to 5 were provided in the DREAMER data set. A classification threshold of 3 was used to mark the labels as positive or negative valence. The ASD classification data sets (ODU and KAU) provided the subjects labeled as ASD (positive) or TD (negative).

The emotions and ASD classification was performed using LSVM, DNN, KNN, DT, and XGB classifiers. The LSVM classifier provided the overall best-case classification results. A subset of only 4 EEG channels was utilized for the classification. The classification result and the four EEG channels subset are summarized in Table 6.1.

Table 6.1: Classification Results of Emotions & ASD Data Sets using EEG Signals.

Data set	Parameter	Channels	Accuracy
DEAP	Valence	FP1, FP2, FC2, T8	90.1%
	Arousal		93.5%
SEED	Valence	T7, T8, FP1, FP2	95.2%
DREAMER	Valence	F3,F4,FC5,T7	98.6%
ODU	ASD	CP2, FC2, F7, FC6	100%
KAU	ASD	F3, T5, O1, O2	95.5%

The classification accuracies of 90.1% and 93.5% were achieved for the valence and arousal classification, respectively, in the DEAP data set. The SEED data set was classified for valence classification with 95.2% accuracy. The DREAMER data set was classified with 98.6% classification respectively for valence classification. The ASD classification was performed with 100% and 95.5% classification accuracy using ODU and KAU data sets respectively.

The selected four EEG channels included FP1, FP2, FC2, and T8 channels for the DEAP data set. The SEED data set provided the best classification results using T7, T8, FP1, and FP2 channels. The selected four channels for the DREAMER data set included F3, F4, FC5, and T7 channels. The four EEG channel subsets for the ODU data set included the CP2, FC2, F7, and FC6 channels. The selected channels for the KAU data set include

F3, T5, O1, and O2 channels.

A set of eight features for the four EEG channels was used for the classification. The eight features are calculated using the brute-force approach from the set of 32 suitable features identified for emotions and ASD classification. The selected features include the fast Fourier transform coefficients, skewness, mean, signal energy, continuous wavelet transform coefficients, auto-correlation, absolute energy, mean absolute change, mean second derivative, change quantiles, energy ratio, and linear trend for emotions and ASD classification. The frequency of the selected features for emotions and ASD classification for each data set is plotted through a wheel diagram in Figure 6-9.

The auto-correlation, energy ratio, linear trend, fast Fourier transform coefficients, skewness and mean features were observed to be highly significant for the ASD classification. Among the selected features for ASD classification, fast Fourier transform coefficients and auto-correlation were observed to be more significant irrespective of the data set. The emotions classification using the DEAP, SEED, and DREAMER data sets provided the maximum classification results using the change quantile, mean absolute change, mean 2nd derivative, absolute energy, fast Fourier transform coefficients, auto-correlation, signal energy, and continuous wavelet transform coefficients. The fast Fourier transform coefficients, mean absolute change, continuous wavelet transform coefficients, and absolute energy was observed to be more significant across multiple data sets for emotions classification as depicted in Figure 6-9.

6.4   Performance Comparison

A brief comparison of the proposed LSVM classifier with selected features and channels with the previous state-of-the-art works using NB, GELM, 3DCANN, RNN, KNN, and CSVM classifiers for emotions and ASD classification is presented in Table 6.2. The classification problem (CLF Prob), Channel count (Ch. Count), best-case percentage accuracy (Acc%), number of validation data sets, and the channel ranking are mentioned in the table.

All the previous works have focused on the emotions or ASD classification and none of them have provided a single framework for both classification problems [24], [26], [42], [89],[115],[46]. This is the first study (to the best of my knowledge) that provides a framework and guideline for both emotions (EMT) and ASD classification. The significance of
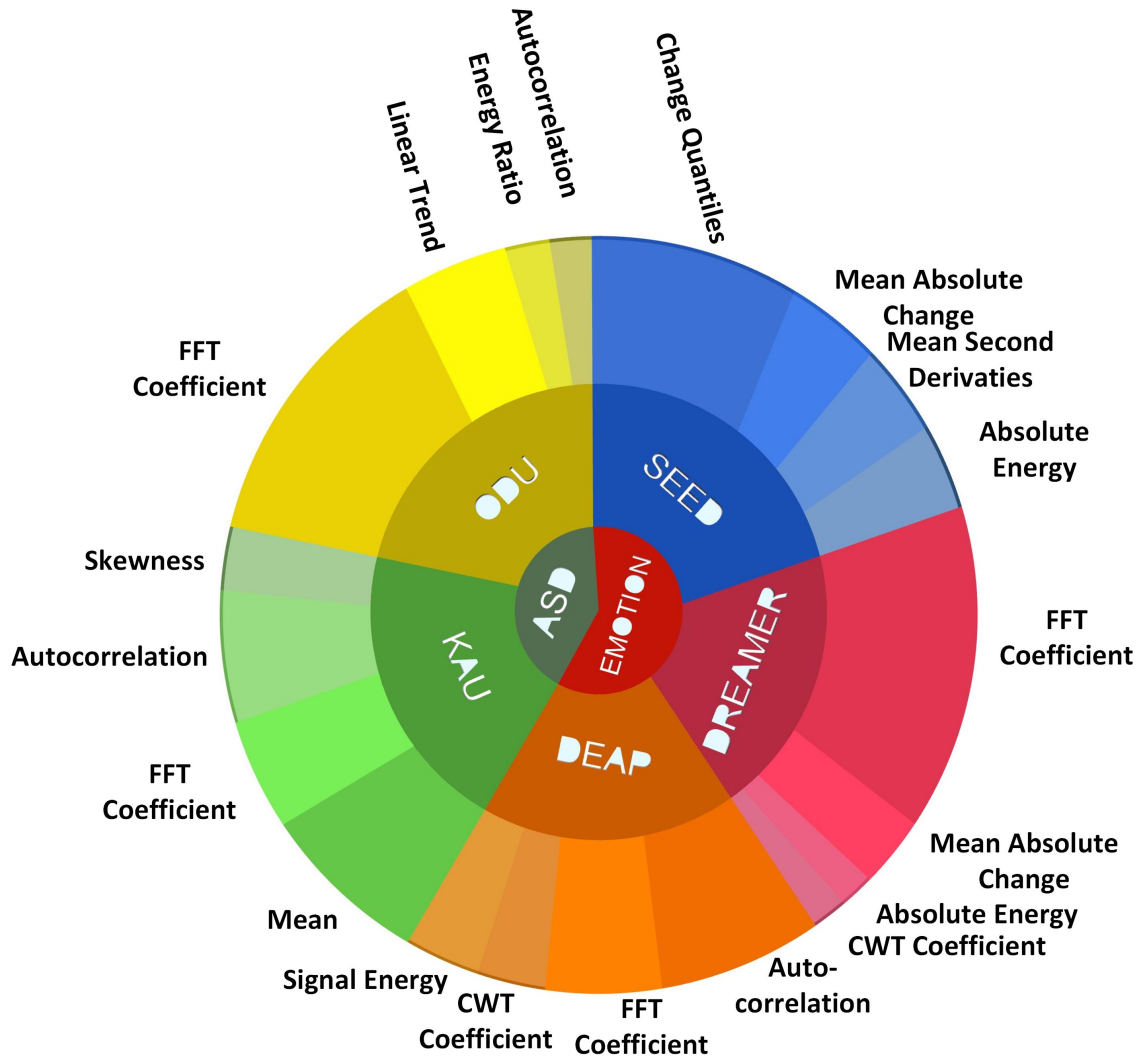
Figure 6-9: Features Wheel Diagram for Emotions & ASD Classification.

EEG-based emotions and ASD classification systems is highly dependent on the channel count, especially for wearable systems [26],[72]. The identification of a minimum (4) number of channels was a primary focus of this study. This study also provided the maximum classification results of 98.6% and 100% classification accuracies for the emotions and ASD classification. The highest number (99) of subjects were analyzed in this study, with the largest number (5) of data sets. This is the only study that provided a detailed analysis for the channel ranking to identify the best four suitable EEG channels.

Table 6.2: Comparison With the State-of-The-Art Works.

	Koelstra	Aslam	Li	Fang	Liu	Aslam	Bouallegue	Alturki	Baygin	T.W
	[24]	[39]	[42]	[26]	[112]	[46]	[113]	[114]	[115]	[66]
CLF Prob	EMT	EMT	EMT	EMT	EMT	ASD	ASD	ASD	ASD	EMT+ASD
Ch. Count	32	8	32	6	62	4	16	16	64	4
Classifier	NB	LSVM	GELM	CNN	3DCANN	LSVM	RNN	KNN	CSVM	LSVM
Acc (%)	62	73.4	88	83.4	97.3	85.5	99.5	98.5	96.4	98.6/100
# of Sub	32	47	74	32	15	12	12	17	54	99
# of Data Set	1	2	3	1	1	1	1	1	1	5
Ch. Rank	No	No	No	No	No	No	No	No	No	Yes

This paper provides a framework for the researchers performing emotions and ASD classification using ML. The most frequently utilized and "benchmark-considered" data sets for the emotions and ASD classification including DEAP, SEED, DREAMER, ODU, and KAU data sets were utilized. A detailed analysis of the data sets including the number of subjects, number of classes per subject, range of label values including valence and arousal, gender and age statistics of each subject, the importance of classification threshold for emotions classification, and the percentage of the positive-negative split in each data set was described.

The location of the EEG channels becomes highly significant for the hardware-based emotions and ASD prediction systems in addition to the channel count. The location of the shortlisted four channels using the 10-20 system is shown in Fig 6-10. It can be observed that temporal and frontal locations are highly significant for emotions prediction, irrespective of the data set. The significance of the prefrontal (FP) region is higher than in other regions and can be observed from DEAP and SEED data set analysis through Fig 6-10 (a),(d). Similarly, the fontal-central and occipital brain regions were observed to be highly significant for the ASD classification.
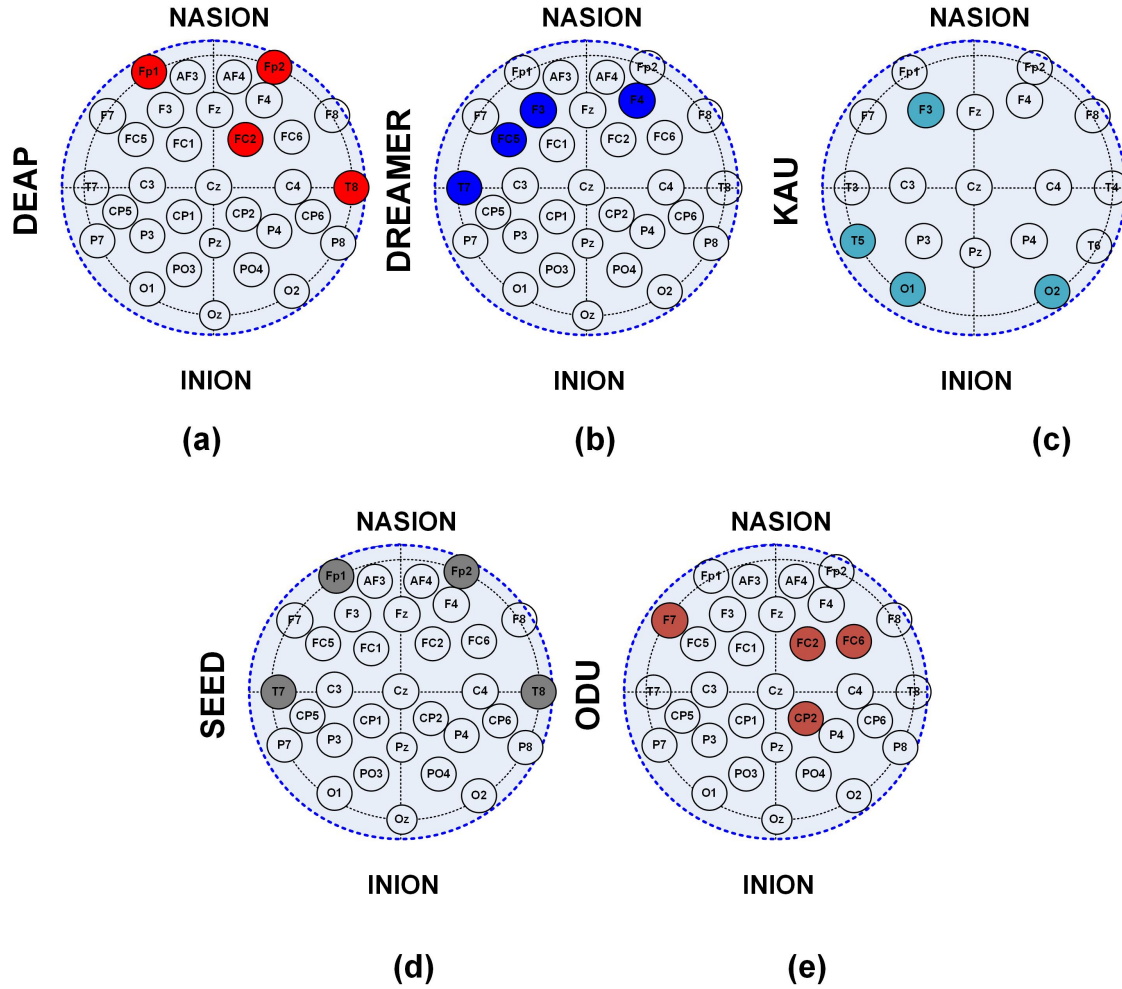
Figure 6-10: Location of 4 Channel EEG subset using 10-20 System on
(a) DEAP (b) SEED (c) DREAMER (d) ODU (e) KAU Data Sets.

6.5 Summary

This chapter has provided a guideline to early researchers for EEG-based emotions and
ASD classification. The following main point are discussed and explained in this chapter:-

- The procedure to quantify and measure human emotion is described.

- Multiple data sets for the emotions and ASD classification are characterized.

- Most frequently used and "benchmark-considered" data sets for emotions and ASD
  prediction are identified.

- The significance of classification threshold in emotions prediction is highlighted.

- The significant brain areas, related EEG channels and significant features for the emotions and ASD prediction are identified using LSFE.

- Performed emotions and ASD prediction using DEAP, SEED, DREAMER, KAU and ODU data sets.

- Achieved excellent classification results for emotions (98.6%) and ASD (100%) prediction.

- The conventional ADOS-2 method for ASD diagnosis and the ADOS-2 scores confirmation with the aid of neurologists is also detailed.

- Utilized a lower number (4) of channels for emotions and ASD prediction.

6.6   Conclusion

In this chapter targeting suitable features and channel identification using LSFE methodology, I have primarily focused on an extensive analysis through LSFE to identify the most suitable channels and brain areas for emotions and ASD classification. I achieved higher classification results than the state-of-the-art using a lower number of channels for emotions and ASD classification. The algorithm was validated on a maximum number of subjects and benchmark data sets. The LSVM classifier used in this work has significantly lower complexity than the other classifiers. The identified brain areas for the emotions classification included the temporal, frontal, and prefrontal regions. The identified regions for the ASD classification included the occipital and frontal-central regions. A subset of the most suitable four EEG channels was identified for each data set which is highly beneficial for the researchers targeting a minimum number of channels. The feature significance highlighted by the extensive LSFE analysis is highly important, especially for the researchers targeting hardware-based systems. The approximated hardware implementations of the highlighted features can be used to develop low-cost (area, power, and energy) hardware systems for emotions and ASD prediction.

Chapter 7

Dissertation Conclusion and Future Work

7.1 Conclusion

This thesis presented the design and development of two wearable DBE processors for ASD children's assistance using their emotions. Irregular emotions and uncontrolled NEOB are major dilemmas for ASD patients causing self-injuries and suicide attempts. The DBE processors are designed and developed using 180 nm CMOS process. The DBE^1 processor was designed and developed using a linear SVM classification algorithm. The DBE^2 processor was designed and developed using a DNN classifier. These processors can be fully integrated with an ULP and low noise AFE for a wearable real-time negative emotions prediction SoC device. The real-time negative emotions prediction can be used to control and regulate negative emotions using suitable biofeedback. Later on, a LSFE analysis was conducted to identify the most suitable channels and features for emotions and ASD prediction.

This PhD thesis presents several contributions, which are summarized below:-

- $\underline{DBE^1}$: The first generation LSVM based processor was designed and developed using a 180 nm CMOS process. The DBE^1 utilizes an area and classification energy of 5.4 mm^2, and $16\mu J/prediction$, respectively. The processor provides the classification accuracy of 73.4%, and was validated using DEEP and SEED data sets [116]. This work was published in the following top international conference, transactions and

book chapters:-

1. **IEEE International Symposium on Circuits and Systems (ISCAS), 2019**

   – **<u>Abdul Rehman Aslam</u>** and Muhammad Awais Bin Altaf, "An 8 channel patient specific neuromorphic processor for the early screening of autistic children through emotion detection," in 2019 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1–5, 2019.

2. **IEEE Transactions on Biomedical Circuits and Systems (TBioCAS), 2020**

   – **<u>Abdul Rehman Aslam</u>** and Muhammad Awais Bin Altaf, "An on-chip processor for chronic neurological disorders assistance using negative affectivity classification," IEEE Transactions on Biomedical Circuits and Systems, vol. 14, no. 4, pp. 838–851, 2020.

3. **Neural Engineering Techniques for Autism Spectrum Disorder, 2021**

   – **<u>Abdul Rehman Aslam</u>** and Muhammad Awais Bin Altaf, "Machine learning–based patient-specific processor for the early intervention in autistic children through emotion detection," in Neural Engineering Techniques for Autism Spectrum Disorder, pp. 287–313, Elsevier, 2021.

- <u>DBE^2</u> : The second generation DNN based processor was designed and developed using a 180 nm CMOS process. The DBE^2 utilizes an area and classification energy of 16 mm^2, and $10.13 \mu J/prediction$, respectively. The processor provides the classification accuracy of 85.4%, and was validated using DEEP and SEED data sets [116]. This work was published in the following top international conference and book chapters:-

1. **IEEE Custom Integrated Circuits Conference (CICC), 2020**

   – **<u>Abdul Rehman Aslam</u>**, Talha Iqbal, Mahnoor Aftab, Wala Saadeh, and Muhammad Awais Bin Altaf, "A10.13uj/classification 2-channel deep neural network-based soc for emotion detection of autistic children," in 2020 IEEE Custom Integrated Circuits Conference (CICC), pp. 1–4, 2020.

2. **IEEE Transactions on Biomedical Circuits and Systems (TBioCAS), 2021**

   – **Abdul Rehman Aslam** and Muhammad Awais Bin Altaf, "A 10.13j/classification 2-channel deep neural network based soc for negative emotion outburst detection of autistic children," IEEE Transactions on Biomedical Circuits and Systems, vol. 15, no. 5, pp. 1039–1052, 2021.

- **LSFE Analysis** : A large-scale feature selection analysis was conducted for best suitable channels and feature selection in emotion and ASD prediction. This work was published in the following top international conference and journal:-

  1. **IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), 2021**

     – **Abdul Rehman Aslam**, Nauman Hafeez, Hadi Heidari, and Muhammad Awais Bin Altaf, "An 8.62 uw processor for autism spectrum disorder classification using shallow neural network," in 2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS), pp. 1–4, 2021.

  2. **Frontiers in Neuroscience (Front. Neurosci), 2022**

     – **Abdul Rehman Aslam**, Nauman Hafeez, Hadi Heidari, and Muhammad Awais Bin Altaf, "Channels and feature identification with large scale feature extraction for emotions and asd classification," Frontiers in Neuroscience, p. 1094, 2022.

7.2   Future Work

I believe that the DBE processors for real-time emotional prediction targeting ASD children would provide promising assistance for ASD patients. However, several areas including binary neural networks, low power AFE, data set collection for ASD patients, and the utility of MMG signals for negative emotions prediction is worth exploring.

Binary Neural Networks

The DNN and deep learning have shown high effectiveness to solve complex machine-learning problems and achieve highly accurate results. Complex DNNs including Convolutional Neural Networks, Long term Short term Memory Networks, Recurrent Neural Networks, Generative Adversarial Networks, RBF networks, Multilayer Perceptrons, Self Organizing Maps, Deep Belief Networks, Auto-encoders, etc. have shown great results for emotions or ASD prediction [117]. But the main bottleneck in their hardware implementations especially wearable SoC devices is the area, power, and energy constraints. A bulky device with a large area with 100% classification results would perish the real objective of miniaturized wearable SoC devices. Similarly, an SoC device with high energy dissipation would not only drain the battery abruptly but also generate too much heat which would make the patients highly uncomfortable.

The binary neural networks are DNNs with binary or 1-bit weights, biases, and activation functions. Binary neural networks have recently shown great potential for complex classification problems. The design and development of a DBE processor using binary neural networks would be highly suitable for low-power and energy SoC devices [118].

Ultra Low Power Analog Front End

The design and development of an ULP AFE is a key element for wearable SoC devices for biomedical applications. The high noise, low signal-to-noise ratio, and high electrode skin impedance are major problems for AFE devices targeting wearable applications. An EEG signal with minimum noise and artifacts is required for high precision emotions prediction or ASD prediction SoC devices.

Electrode Skin Impedance impedes the EEG signal generated by the surface between the skin contact and EEG electrodes. High electrode skin impedance is a major noise factor in the EEG signals. The design and development of an AFE with minimum noise factor, minimum electrode skin impedance and ULP is required to be integrated with a high precision DBE processor for a fully integrated wearable SoC device for emotions prediction, ASD prediction, or other similar biomedical applications.

Physiological Signals Data Bank for ASD Patients

Limited physiological signals are a major challenge faced by researchers targeting the design and development of solutions for the diagnosis and assistance of ASD patients. The sensitive nature of ASD patients and the requirements to fulfill the ethical approvals are major barriers in the data sets collection. I would target the collection of physiological (EEG) signal data set of ASD patients after necessary ethical approvals with the aid of neurologists.

Mechanomyogram Signals for Emotions Prediction

MMG signals are mechanical signals generated by the muscles' contraction which are measured through mechanical vibrations on the skin surface [119]. It is recently investigated that the MMG signals can be used to detect human emotions using forearm MMG sensors [120]. Since MMG measures mechanical signals, it has several benefits over electrical signals including chronic implants and environmental effects [121]. I, therefore, aim to explore the utilization of MMG signals for emotional prediction SoC.

Bibliography

[1] Brigham and women's hospital, "Intracranial electrical stimulation testing (ecs) and intracranial eeg," *[Online].* https://golbylab.bwh.harvard.edu/intracranial-electrical-stimulation-testing-ecs-and-intracranial-eeg/.

[2] Brigham and women's hospital, "London's eeg, neurofeedback and brain stimulation centre," *[Online].* https://www.brightbraincentre.co.uk/electroencephalogram-eeg-brainwaves/.

[3] F. Happé and U. Frith, "The weak coherence account: detail-focused cognitive style in autism spectrum disorders," *Journal of autism and developmental disorders*, vol. 36, no. 1, pp. 5–25, 2006.

[4] M. R. Herbert *et al.*, "Autism: a brain disorder or a disorder that affects the brain," *Clinical Neuropsychiatry*, vol. 2, no. 6, pp. 354–379, 2005.

[5] J. Xiong, S. Chen, N. Pang, X. Deng, L. Yang, F. He, L. Wu, C. Chen, F. Yin, and J. Peng, "Neurological diseases with autism spectrum disorder: role of asd risk genes," *Frontiers in neuroscience*, vol. 13, p. 349, 2019.

[6] C. for Disease Control and Prevention, "Data and statistics on autism spectrum disorder," https://www.cdc.gov/ncbddd/autism/data.html.

[7] M. S. Anwar, M. Tahir, K. Nusrat, and M. R. Khan, "Knowledge, awareness, and perceptions regarding autism among parents in karachi, pakistan," *Cureus*, vol. 10, no. 9, 2018.

[8] Tribune, "Autism in pakistan," *[Online].* https://tribune.com.pk/story/2108977/6-autism-in-pakistan.

[9] M. Rutter, "Aetiology of autism: findings and questions," *Journal of Intellectual Disability Research*, vol. 49, no. 4, pp. 231–238, 2005.

[10] M. Blaxill, T. Rogers, and C. Nevison, "Autism tsunami: the impact of rising prevalence on the societal cost of autism in the united states," *Journal of Autism and Developmental Disorders*, vol. 52, no. 6, pp. 2627–2643, 2022.

[11] L. Hewitson, J. A. Mathews, M. Devlin, C. Schutte, J. Lee, and D. C. German, "Blood biomarker discovery for autism spectrum disorder: A proteomic analysis," *PloS one*, vol. 16, no. 2, p. e0246581, 2021.

[12] E. Rydzewska, L. A. Hughes-McCormack, C. Gillberg, A. Henderson, C. MacIntyre, J. Rintoul, and S.-A. Cooper, "Age at identification, prevalence and general health

of children with autism: observational study of a whole country population," *BMJ open*, vol. 9, no. 7, p. e025904, 2019.

[13] A. McCrimmon and K. Rostad, "Test review: Autism diagnostic observation schedule, (ados-2) manual (part ii): Toddler module," 2014.

[14] S. H. Park, Y. J. C. Song, E. A. Demetriou, K. L. Pepper, A. Norton, E. E. Thomas, I. B. Hickie, D. F. Hermens, N. Glozier, and A. J. Guastella, "Disability, functioning, and quality of life among treatment-seeking young autistic adults and its relation to depression, anxiety, and stress," *Autism*, vol. 23, no. 7, pp. 1675–1686, 2019.

[15] A. C. Samson, J. M. Phillips, K. J. Parker, S. Shah, J. J. Gross, and A. Y. Hardan, "Emotion dysregulation and the core features of autism spectrum disorder," *Journal of Autism and developmental Disorders*, vol. 44, no. 7, pp. 1766–1772, 2014.

[16] H. Kassinove and D. G. Sukhodolsky, "Anger disorders: Basic science and practice issues," *Issues in comprehensive pediatric nursing*, vol. 18, no. 3, pp. 173–205, 1995.

[17] M. J. Hollocks, J. W. Lerh, I. Magiati, R. Meiser-Stedman, and T. S. Brugha, "Anxiety and depression in adults with autism spectrum disorder: a systematic review and meta-analysis," *Psychological medicine*, vol. 49, no. 4, pp. 559–572, 2019.

[18] S. D. Mayes, A. A. Gorman, J. Hillwig-Garcia, and E. Syed, "Suicide ideation and attempts in children with autism," *Research in Autism Spectrum Disorders*, vol. 7, no. 1, pp. 109–119, 2013.

[19] S. Ryan, "'meltdowns', surveillance and managing emotions; going out with children with autism," *Health & place*, vol. 16, no. 5, pp. 868–875, 2010.

[20] S. Sinha, R. A. McGovern, and S. A. Sheth, "Deep brain stimulation for severe autism: from pathophysiology to procedure," *Neurosurgical focus*, vol. 38, no. 6, p. E3, 2015.

[21] H. Yan, L. Siegel, S. Breitbart, C. Gorodetsky, A. Fasano, A. Rahim, A. Loh, A. V. Kulkarni, and G. M. Ibrahim, "An open-label prospective pilot trial of nucleus accumbens deep brain stimulation for children with autism spectrum disorder and severe, refractory self-injurious behavior: study protocol," *Pilot and Feasibility Studies*, vol. 8, no. 1, pp. 1–10, 2022.

[22] C. A. Mazefsky, J. Herrington, M. Siegel, A. Scarpa, B. B. Maddox, L. Scahill, and S. W. White, "The role of emotion regulation in autism spectrum disorder," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 52, no. 7, pp. 679–688, 2013.

[23] O. Valentin, M. Ducharme, G. Crétot-Richert, H. Monsarrat-Chanon, G. Viallet, A. Delnavaz, and J. Voix, "Validation and benchmarking of a wearable eeg acquisition platform for real-world applications," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 1, pp. 103–111, 2019.

[24] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 18–31, 2011.

[25] H. A. Gonzalez, S. Muzaffar, J. Yoo, and I. M. Elfadel, "Biocnn: A hardware inference engine for eeg-based emotion detection," *IEEE Access*, vol. 8, pp. 140896–140914, 2020.

[26] W.-C. Fang, K.-Y. Wang, N. Fahier, Y.-L. Ho, and Y.-D. Huang, "Development and validation of an eeg-based real-time emotion recognition system using edge ai computing platform with convolutional neural network system-on-chip design," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 4, pp. 645–657, 2019.

[27] M. A. Bin Altaf and J. Yoo, "A 1.83 μj/classification, 8-channel, patient-specific epileptic seizure classification soc using a non-linear support vector machine," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 10, no. 1, pp. 49–60, 2016.

[28] C. E. Izard, *Human emotions*. Springer Science & Business Media, 2013.

[29] P. Ekman, "Basic emotions," *Handbook of cognition and emotion*, vol. 98, no. 45-60, p. 16, 1999.

[30] R. Plutchik, "A general psychoevolutionary theory of emotion," pp. 3–33, 1980.

[31] W. G. Parrott, *Emotions in social psychology: Essential readings*. psychology press, 2001.

[32] S. S. Tomkins, "Affect as amplification: Some modifications in theory," in *Theories of emotion*, pp. 141–164, Elsevier, 1980.

[33] T. Eerola and J. K. Vuoskoski, "A comparison of the discrete and dimensional models of emotion in music," *Psychology of Music*, vol. 39, no. 1, pp. 18–49, 2011.

[34] J. A. Russell, "A circumplex model of affect.," *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[35] V. Jurcak, D. Tsuzuki, and I. Dan, "10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems," *Neuroimage*, vol. 34, no. 4, pp. 1600–1611, 2007.

[36] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[37] R. Y. Cai, A. L. Richdale, M. Uljarević, C. Dissanayake, and A. C. Samson, "Emotion regulation in autism spectrum disorder: Where we are and where we need to go," *Autism Research*, vol. 11, no. 7, pp. 962–978, 2018.

[38] A. R. Aslam and M. A. B. Altaf, "An on-chip processor for chronic neurological disorders assistance using negative affectivity classification," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 14, no. 4, pp. 838–851, 2020.

[39] A. R. Aslam and M. A. B. Altaf, "A 10.13µj/classification 2-channel deep neural network based soc for negative emotion outburst detection of autistic children," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 15, no. 5, pp. 1039–1052, 2021.

[40] L. Lu, J. Zhang, Y. Xie, F. Gao, S. Xu, X. Wu, Z. Ye, *et al.*, "Wearable health devices in health care: narrative systematic review," *JMIR mHealth and uHealth*, vol. 8, no. 11, p. e18907, 2020.

[41] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from eeg using higher order crossings," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 186–197, 2010.

[42] P. Li, H. Liu, Y. Si, C. Li, F. Li, X. Zhu, X. Huang, Y. Zeng, D. Yao, Y. Zhang, *et al.*, "Eeg based emotion recognition by combining functional connectivity network and local activations," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 10, pp. 2869–2881, 2019.

[43] W. Li, Z. Zhang, B. Hou, and X. Li, "A novel spatio-temporal field for emotion recognition based on eeg signals," *IEEE Sensors Journal*, vol. 21, no. 23, pp. 26941–26950, 2021.

[44] L. S. Mokatren, R. Ansari, A. E. Cetin, A. D. Leow, O. A. Ajilore, H. Klumpp, and F. T. Yarman Vural, "Eeg classification by factoring in sensor spatial configuration," *IEEE Access*, vol. 9, pp. 19053–19065, 2021.

[45] T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, and Z. Cui, "Mped: A multi-modal physiological emotion database for discrete emotion recognition," *IEEE Access*, vol. 7, pp. 12177–12191, 2019.

[46] S. Liu, X. Wang, L. Zhao, B. Li, W. Hu, J. Yu, and Y. Zhang, "3dcann: A spatio-temporal convolution attention neural network for eeg emotion recognition," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[47] F. Demir, N. Sobahi, S. Siuly, and A. Sengur, "Exploring deep learning features for automatic classification of human emotion using eeg rhythms," *IEEE Sensors Journal*, vol. 21, no. 13, pp. 14923–14930, 2021.

[48] G. Chen, X. Zhang, Y. Sun, and J. Zhang, "Emotion feature analysis and recognition based on reconstructed eeg sources," *IEEE Access*, vol. 8, pp. 11907–11916, 2020.

[49] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, "Analysis of eeg signals and facial expressions for continuous emotion detection," *IEEE Transactions on Affective Computing*, vol. 7, no. 1, pp. 17–28, 2016.

[50] Z. Gao, X. Wang, Y. Yang, Y. Li, K. Ma, and G. Chen, "A channel-fused dense convolutional network for eeg-based emotion recognition," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, no. 4, pp. 945–954, 2021.

[51] Y. Luo, Q. Fu, J. Xie, Y. Qin, G. Wu, J. Liu, F. Jiang, Y. Cao, and X. Ding, "Eeg-based emotion classification using spiking neural networks," *IEEE Access*, vol. 8, pp. 46007–46016, 2020.

[52] Y.-D. Huang, K.-Y. Wang, Y.-L. Ho, C.-Y. He, and W.-C. Fang, "An edge ai system-on-chip design with customized convolutional-neural-network architecture for real-time eeg-based affective computing system," in *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 1–4, 2019.

[53] C.-J. Yang, W.-C. Li, M.-T. Wan, and W.-C. Fang, "Real-time eeg-based affective computing using on-chip learning long-term recurrent convolutional network," in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, 2021.

[54] K.-Y. Wang, Y.-L. Ho, Y.-D. Huang, and W.-C. Fang, "Design of intelligent eeg system for human emotion recognition with convolutional neural network," in *2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, pp. 142–145, 2019.

[55] C. Ardito, I. Bortone, T. Colafiglio, T. D. Noia, E. D. Sciascio, D. Lofù, F. Narducci, R. Sardone, and P. Sorino, "Brain computer interface: Deep learning approach to predict human emotion recognition," in *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2689–2694, 2022.

[56] X. Shen, X. Liu, X. Hu, D. Zhang, and S. Song, "Contrastive learning of subject-invariant eeg representations for cross-subject emotion recognition," *IEEE Transactions on Affective Computing*, pp. 1–1, 2022.

[57] J. W. Li, S. Barma, P. U. Mak, F. Chen, C. Li, M. T. Li, M. I. Vai, and S. H. Pun, "Single-channel selection for eeg-based emotion recognition using brain rhythm sequencing," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 6, pp. 2493–2503, 2022.

[58] A. Menon, D. Sun, S. Sabouri, K. Lee, M. Aristio, H. Liew, and J. M. Rabaey, "A highly energy-efficient hyperdimensional computing processor for biosignal classification," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 16, no. 4, pp. 524–534, 2022.

[59] S. M. Alarcão and M. J. Fonseca, "Emotions recognition using eeg signals: A survey," *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 374–393, 2019.

[60] B. S. Mashford, A. J. Yepes, I. Kiral-Kornek, J. Tang, and S. Harrer, "Neural-network-based analysis of eeg data using the neuromorphic truenorth chip for brain-machine interfaces," *IBM Journal of Research and Development*, vol. 61, no. 2/3, pp. 7–1, 2017.

[61] W. Saadeh, M. A. Bin Altaf, and S. A. Butt, "A wearable neuro-degenerative diseases detection system based on gait dynamics," in *2017 IFIP/IEEE International Conference on Very Large Scale Integration (VLSI-SoC)*, pp. 1–6, 2017.

[62] H. A. Gonzalez, R. M. George, S. Muzaffar, J. Acevedo, S. Hoeppner, C. Mayr, J. Yoo, F. H. Fitzek, and I. Elfadel, "Hardware acceleration of eeg-based emotion classification systems: A comprehensive survey," *IEEE Transactions on Biomedical Circuits and Systems*, 2021.

[63] C. E. Izard, "Emotion theory and research: Highlights, unanswered questions, and emerging issues," *Annual review of psychology*, vol. 60, p. 1, 2009.

[64] L. F. Barrett, B. Mesquita, K. N. Ochsner, and J. J. Gross, "The experience of emotion," *Annual review of psychology*, vol. 58, p. 373, 2007.

[65] L. F. Barrett, "Discrete emotions or dimensions? the role of valence focus and arousal focus," *Cognition & Emotion*, vol. 12, no. 4, pp. 579–599, 1998.

[66] A. R. Aslam, N. Hafeez, H. Heidari, and M. A. B. Altaf, "Channels and feature identification with large scale feature extraction for emotions and asd classification," *Frontiers in Neuroscience*, p. 1094, 2022.

[67] M. A. Bin Altaf, C. Zhang, and J. Yoo, "A 16-channel patient-specific seizure onset and termination detection soc with impedance-adaptive transcranial electrical stimulator," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 11, pp. 2728–2740, 2015.

[68] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying stable patterns over time for emotion recognition from eeg," *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 417–429, 2019.

[69] A. Topic, M. Russo, M. Stella, and M. Saric, "Emotion recognition using a reduced set of eeg channels based on holographic feature maps," *Sensors*, vol. 22, no. 9, p. 3248, 2022.

[70] J. Wang, J. Barstein, L. E. Ethridge, M. W. Mosconi, Y. Takarae, and J. A. Sweeney, "Resting state eeg abnormalities in autism spectrum disorders," *Journal of neurodevelopmental disorders*, vol. 5, no. 1, pp. 1–14, 2013.

[71] D. M. Muñoz, D. F. Sanchez, C. H. Llanos, and M. Ayala-Rincón, "Tradeoff of fpga design of a floating-point library for arithmetic operators," *Journal of Integrated Circuits and Systems*, vol. 5, no. 1, pp. 42–52, 2010.

[72] A. R. Aslam and M. A. B. Altaf, "An 8 channel patient specific neuromorphic processor for the early screening of autistic children through emotion detection," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, 2019.

[73] J. M. Bland and D. G. Altman, "Statistics notes. logarithms.," *BMJ: British Medical Journal*, vol. 312, no. 7032, p. 700, 1996.

[74] D. Singh and B. Singh, "Feature wise normalization: An effective way of normalizing data," *Pattern Recognition*, vol. 122, p. 108307, 2022.

[75] T. Kavzoglu and I. Colkesen, "A kernel functions analysis for support vector machines for land cover classification," *International Journal of Applied Earth Observation and Geoinformation*, vol. 11, no. 5, pp. 352–359, 2009.

[76] M. Hofmann, "Support vector machines-kernels and the kernel trick," *Notes*, vol. 26, no. 3, pp. 1–16, 2006.

[77] K. H. Lee, S.-Y. Kung, and N. Verma, "Low-energy formulations of support vector machine kernel functions for biomedical sensor applications," *Journal of Signal Processing Systems*, vol. 69, no. 3, pp. 339–349, 2012.

[78] D. Lewis, "Complex logarithmic number system arithmetic using high-radix redundant cordic algorithms," in *Proceedings 14th IEEE Symposium on Computer Arithmetic (Cat. No. 99CB36336)*, pp. 194–203, IEEE, 1999.

[79] H. Ullah, M. Uzair, A. Mahmood, M. Ullah, S. D. Khan, and F. A. Cheikh, "Internal emotion classification using eeg signal with sparse discriminative ensemble," *IEEE Access*, vol. 7, pp. 40144–40153, 2019.

[80] Y. Peng, S. Wang, X. Long, and B.-L. Lu, "Discriminative graph regularized extreme learning machine and its application to face recognition," *Neurocomputing*, vol. 149, pp. 340–353, 2015.

[81] L. Pessoa and R. Adolphs, "Emotion processing and the amygdala: from a'low road'to'many roads' of evaluating biological significance," *Nature reviews neuroscience*, vol. 11, no. 11, pp. 773–782, 2010.

[82] G. Li, M.-H. Chen, G. Li, D. Wu, C. Lian, Q. Sun, R. J. Rushmore, and L. Wang, "Volumetric analysis of amygdala and hippocampal subfields for infants with autism," *Journal of Autism and Developmental Disorders*, pp. 1–15, 2022.

[83] A. Weinberg and E. D. Klonsky, "The effects of self-injury on acute negative arousal: A laboratory simulation," *Motivation and Emotion*, vol. 36, no. 2, pp. 242–254, 2012.

[84] E. D. Klonsky, "The functions of deliberate self-injury: A review of the evidence," *Clinical psychology review*, vol. 27, no. 2, pp. 226–239, 2007.

[85] J. A. Coan and J. J. Allen, "Frontal eeg asymmetry as a moderator and mediator of emotion," *Biological psychology*, vol. 67, no. 1-2, pp. 7–50, 2004.

[86] S. J. Johnston, S. G. Boehm, D. Healy, R. Goebel, and D. E. Linden, "Neurofeedback: A promising tool for the self-regulation of emotion networks," *Neuroimage*, vol. 49, no. 1, pp. 1066–1072, 2010.

[87] H. Chang, Y. Zong, W. Zheng, C. Tang, J. Zhu, and X. Li, "Depression assessment method: An eeg emotion recognition framework based on spatiotemporal neural network," *Frontiers in Psychiatry*, vol. 12, 2021.

[88] T. Zhan, S. Guraya, and H. Kassiri, "A resource-optimized vlsi architecture for patient-specific seizure detection using frontal-lobe eeg," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, 2019.

[89] A. R. Aslam, T. Iqbal, M. Aftab, W. Saadeh, and M. A. Bin Altaf, "A10.13uj/classification 2-channel deep neural network-based soc for emotion detection of autistic children," in *2020 IEEE Custom Integrated Circuits Conference (CICC)*, pp. 1–4, 2020.

[90] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis," *IEEE Transactions on Affective Computing*, vol. 1, no. 2, pp. 81–97, 2010.

[91] D. D. Chakladar and S. Chakraborty, "Eeg based emotion classification using "correlation based subset selection"," *Biologically inspired cognitive architectures*, vol. 24, pp. 98–106, 2018.

[92] H. Mohsen, E.-S. A. El-Dahshan, E.-S. M. El-Horbaty, and A.-B. M. Salem, "Classification using deep learning neural networks for brain tumors," *Future Computing and Informatics Journal*, vol. 3, no. 1, pp. 68–71, 2018.

[93] J. Feng and S. Lu, "Performance analysis of various activation functions in artificial neural networks," in *Journal of physics: conference series*, vol. 1237, p. 022030, IOP Publishing, 2019.

[94] J. Kamruzzaman and S. M. Aziz, "A note on activation function in multilayer feedforward learning," in *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*, vol. 1, pp. 519–523, IEEE, 2002.

[95] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, "Efficient processing of deep neural networks: A tutorial and survey," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2295–2329, 2017.

[96] J. Lee, C. Kim, S. Kang, D. Shin, S. Kim, and H.-J. Yoo, "Unpu: An energy-efficient deep neural network accelerator with fully variable weight bit precision," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 1, pp. 173–185, 2019.

[97] H. A. Gonzalez, J. Yoo, and I. M. Elfadel, "Eeg-based emotion detection using unsupervised transfer learning," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 694–697, 2019.

[98] X. Jiang, G.-B. Bian, and Z. Tian, "Removal of artifacts from eeg signals: a review," *Sensors*, vol. 19, no. 5, p. 987, 2019.

[99] S. Katsigiannis and N. Ramzan, "Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 98–107, 2018.

[100] N. Watt, A. M. Wetherby, A. Barber, and L. Morgan, "Repetitive and stereotyped behaviors in children with autism spectrum disorders in the second year of life," *Journal of autism and developmental disorders*, vol. 38, no. 8, pp. 1518–1533, 2008.

[101] Y. Jayawardana, M. Jaime, and S. Jayarathna, "Analysis of temporal relationships between asd and brain activity through eeg and machine learning," in *2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI)*, pp. 151–158, IEEE, 2019.

[102] M. J. Alhaddad, M. I. Kamel, H. M. Malibary, E. A. Alsaggaf, K. Thabit, F. Dahlwi, and A. A. Hadi, "Diagnosis autism by fisher linear discriminant analysis flda via eeg," *International Journal of Bio-Science and Bio-Technology*, vol. 4, no. 2, pp. 45–54, 2012.

[103] B. D. Fulcher, "Feature-based time-series analysis," *arXiv preprint arXiv:1709.08055*, 2017.

[104] M. Barandas, D. Folgado, L. Fernandes, S. Santos, M. Abreu, P. Bota, H. Liu, T. Schultz, and H. Gamboa, "Tsfel: Time series feature extraction library," *SoftwareX*, vol. 11, p. 100456, 2020.

[105] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package)," *Neurocomputing*, vol. 307, pp. 72–77, 2018.

[106] M. Christ, "Overview on extracted features." `https://tsfresh.readthedocs.io/en/latest/text/list_of_features.html`, 2016. [Online].

[107] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[108] J. Gómez-Ramírez, M. Ávila-Villanueva, and M. Á. Fernández-Blázquez, "Selecting the most important self-assessed features for predicting conversion to mild cognitive impairment with random forest and permutation-based methods," *Scientific reports*, vol. 10, no. 1, pp. 1–15, 2020.

[109] X. Tong, D. An, F. Xiao, D. Lei, R. Niu, W. Li, J. Ren, W. Liu, Y. Tang, L. Zhang, *et al.*, "Real-time effects of interictal spikes on hippocampus and amygdala functional connectivity in unilateral temporal lobe epilepsy: An eeg-fmri study," *Epilepsia*, vol. 60, no. 2, pp. 246–254, 2019.

[110] N. S. Suhaimi, J. Mountstephens, and J. Teo, "Eeg-based emotion recognition: A state-of-the-art review of current trends and opportunities," *Computational intelligence and neuroscience*, vol. 2020, 2020.

[111] A. Abdiansah and R. Wardoyo, "Time complexity analysis of support vector machines (svm) in libsvm," *International journal computer and application*, vol. 128, no. 3, pp. 28–34, 2015.

[112] A. R. Aslam, N. Hafeez, H. Heidari, and M. A. B. Altaf, "An 8.62 w processor for autism spectrum disorder classification using shallow neural network," in *2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, pp. 1–4, 2021.

[113] G. Bouallegue, R. Djemal, S. A. Alshebeili, and H. Aldhalaan, "A dynamic filtering df-rnn deep-learning-based approach for eeg-based neurological disorders diagnosis," *IEEE Access*, vol. 8, pp. 206992–207007, 2020.

[114] F. A. Alturki, M. Aljalal, A. M. Abdurraqeeb, K. Alsharabi, and A. A. Al-Shamma'a, "Common spatial pattern technique with eeg signals for diagnosis of autism and epilepsy disorders," *IEEE Access*, vol. 9, pp. 24334–24349, 2021.

[115] M. Baygin, S. Dogan, T. Tuncer, P. D. Barua, O. Faust, N. Arunkumar, E. W. Abdulhay, E. E. Palmer, and U. R. Acharya, "Automated asd detection using hybrid deep lightweight features extracted from eeg signals," *Computers in Biology and Medicine*, vol. 134, p. 104548, 2021.

[116] A. R. Aslam and M. A. B. Altaf, "Machine learning–based patient-specific processor for the early intervention in autistic children through emotion detection," in *Neural Engineering Techniques for Autism Spectrum Disorder*, pp. 287–313, Elsevier, 2021.

[117] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE access*, vol. 7, pp. 53040–53065, 2019.

[118] C. Yuan and S. S. Agaian, "A comprehensive review of binary neural network," *arXiv preprint arXiv:2110.06804*, 2021.

[119] M. Watakabe, K. Mita, K. Akataki, and K. Ito, "Reliability of the mechanomyogram detected with an accelerometer during voluntary contractions," *Medical and Biological Engineering and Computing*, vol. 41, no. 2, pp. 198–202, 2003.

[120] B. Zhou, T. Ghose, and P. Lukowicz, "Expressure: detect expressions related to emotional and cognitive activities using forehead textile pressure mechanomyography," *Sensors*, vol. 20, no. 3, p. 730, 2020.

[121] M. T. Tarata, "Mechanomyography versus electromyography, in monitoring the muscular fatigue," *Biomedical engineering online*, vol. 2, no. 1, pp. 1–10, 2003.