



Aslam, A. R., Hafeez, N., Heidari, H. and Altaf, M. A. B. (2021) An 8.62  $\mu$ W Processor for Autism Spectrum Disorder Classification using Shallow Neural Network. In: 3rd IEEE International Conference on Artificial Intelligence Circuits & Systems, 6-9 Jun 2021, ISBN 9781665430258 (doi:[10.1109/AICAS51828.2021.9458412](https://doi.org/10.1109/AICAS51828.2021.9458412))

The material cannot be used for any other purpose without further permission of the publisher and is for private use only.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/238162/>

Deposited on 06 April 2021

Enlighten – Research publications by members of the University of  
Glasgow  
<http://eprints.gla.ac.uk>

# An 8.62 $\mu$ W Processor for Autism Spectrum Disorder Classification using Shallow Neural Network

Abdul Rehman Aslam  
Electrical Engineering  
Lahore University of Management  
Sciences (LUMS)  
Lahore, Pakistan  
[17060056@lums.edu.pk](mailto:17060056@lums.edu.pk)

Muhammad Awais Bin Altaf  
Electrical Engineering  
Lahore University of Management  
Sciences (LUMS)  
Lahore, Pakistan  
[awais.altaf@lums.edu.pk](mailto:awais.altaf@lums.edu.pk)

Nauman Hafeez  
Mechanical & Aerospace Engineering  
Brunel University London  
London, United Kingdom  
[nauman.hafeez@brunel.ac.uk](mailto:nauman.hafeez@brunel.ac.uk)

Hadi Heidari  
Electronic and Nanoscale Engineering  
University of Glasgow  
Scotland, United Kingdom  
[hadi.heidari@glasgow.ac.uk](mailto:hadi.heidari@glasgow.ac.uk)

**Abstract**— Autism Spectrum Disorder (ASD) is the most prevalent child neurological and developmental disorder causing cognitive and behavioral impairments. The early diagnosis is an urgent need for the treatment and rehabilitation of ASD patients. This work presents electroencephalogram (EEG) based ASD classification processor implementation that targets a patch-form factor sensor for long time monitoring in a wearable device. A patient is classified as ASD or typically developing using scalp EEG. The selection of frontal and parietal lobe electrodes causes minimum uneasiness to the children. The proposed and implemented algorithm utilizes only four EEG electrodes. The processor is implemented and validated on Artix-7 FPGA, requiring only 26229 lookup tables and 15180 flip flops. The hardware efficient implementation of the complex kurtosis value and Katz fractal dimension features using kurtosis value indicator and Katz fractal dimension indicator with 54% and 38% efficient implementations, respectively, is provided. A hardware feasible shallow neural network architecture is used for the ASD classification. The implemented system classifies the ASD with a high classification accuracy of 85.5% using the power and latency of 8.62  $\mu$ W and 2.25 milliseconds, respectively.

**Keywords**— Autism Spectrum Disorder, System on Chip, Wearable devices)

## I. INTRODUCTION

Autism spectrum disorder (ASD) is a wide spectrum of neurological disorders including genetic and non-genetic factors. The term “spectrum” represents the widespread series of impairments associated with the disorder, causing the early diagnosis challenging. The recent estimates show a significant rise in the number of ASD patients across all ethnicities and socio economic ranks [1].

The ASD diagnosis is standardly performed by Autism diagnostic observation schedule, 2nd Edition (ADOS-2) [1], requiring extensive and frequent behavioural observations leading to late diagnosis [1]. The ADOS-2 involves a comprehensive analysis and observations of communication (CSC) scores, social interaction (SCI) scores, imagination, and creativity (IMC) score, stereotyped behaviours (STB) score, and their comparisons with a cut off values table. The medical practitioners/ neurologists then evaluate the children as ASD or typically developing (TD). These evaluations take ample time and may be avoided by many parents due to the

feeling of disgrace and repeated visits to neurologists. Fig. 1 shows the difference between the conventional ADOS-2 diagnosis and the proposed solution to solve the pain problem. The proposed solution (longer-term) would be able to diagnose a child as ASD or TD earlier in the form of a wearable head-band system on chip (SoC) processor. The processor would pre-process the brain waves (EEG), extract suitable features, and classify the child as ASD or TD using suitable machine learning (ML) or deep neural network (DNN) classification method with limited electrodes. It would not only avoid the stigma associated with the prolonged diagnosis but also reduce the rehabilitation costs due to early intervention [4].

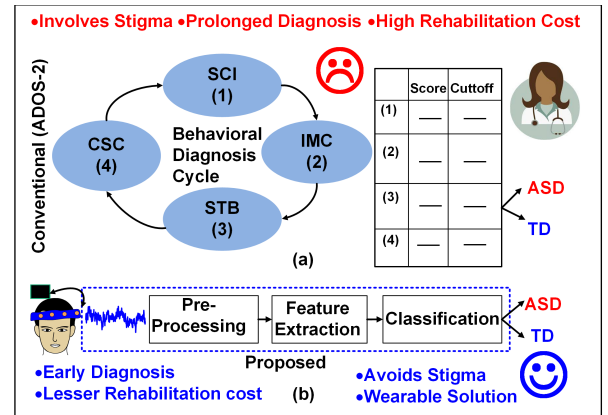


Fig. 1. a) Conventional ADOS-2 diagnosis (b) Proposed solution

One of the biggest challenges in the ASD detection at an early stage is uncooperative behaviour of ASD children, therefore we are proposing to develop a miniaturized wearable device to record and process EEG data. Only transmitting the EEG data wirelessly for remote processing will consume  $>15mW$  power, which is not suitable for children under the age of 4 years [5]. Hence, a fully on-chip low power system ( $\sim 0.5mW$ ) is necessitated to extract the features and classify the ASD from EEG data on the sensor to assist the neurologist in the early detection.

Electroencephalogram (EEG) signals record the electrical activity inside the human brain using a certain number of electrodes. Despite the various challenges related to EEG signal acquisition including noise and artifacts, there is

significant research to show the effectivity of scalp EEG for ASD diagnosis [6]. There are some solutions to assist ASD children using their emotions [7]- [8]. But no hardware-based ASD prediction processor is available. This paper presents the first hardware-based low-power processor to classify ASD patients using a dataset recorded, trained, and tested on ADOS-2 confirmed ASD patients.

## II. METHODS AND TECHNIQUES

The ASD prediction using the EEG signal involves the acquisition of EEG data, EEG signal pre-processing for noise and artifacts removal, suitable feature extraction, and ML/deep learning classification.

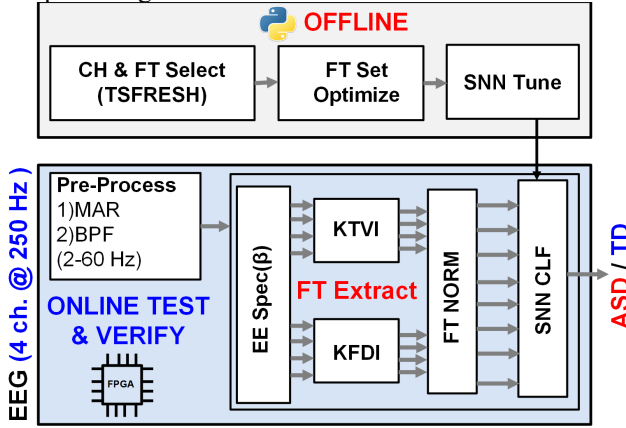


Fig. 2. Top-level block diagram of the ASD classification processor

Fig. 2 shows the top-level diagram of our ASD classification processor. The top part of the figure shows the offline analysis carried out in different python packages to identify the suitable features and channels required for ASD classification [10]. The selected features set and channels list were further optimized for the hardware realizable implementation and a shallow neural network (SNN) classifier was trained offline. Neural networks are capable of learning small datasets quickly [11]. The SNN architecture presented in this paper provided us better classification results with minimum hardware resources against other machine learning models. The SNN parameters (weights and biases) were then uploaded to our ASD classification processor for online testing and verification. The bottom part of the figure shows our hardware-based ASD classification processor including the EEG pre-processing unit, feature extraction unit, and SNN classifier. The processor was implemented using Xilinx Artix-7 FPGA. The processor classifies a subject as ASD or TD using the EEG signals of the selected four channels.

### A. Dataset

To develop an efficient algorithm for the ASD classification, we have utilized the data recorded by Y. Jayawardana et.al [6]. The data provides the EEG dataset of 17 participants including 8 ASD patients and 9 TD subjects using 32 electrodes. The EEG data was sampled at 250 Hz for 9 minutes duration.

### B. Channels Selection

The selection of a limited number of channels and their scalp locations are quite important for the continuous monitoring of EEG data of ASD patients [7]. Due to the discomfort involved and the hardware infeasibility, a large

number of EEG channels and bulky EEG headsets are not suitable for a wearable device. The initial analysis of the EEG signals for ASD and TD subjects identified different channels and features to be quite important for ASD classification. A four channel set (F7, F8, CP5, CP2) was chosen for our ASD classification processor. These channels differentiate ASD and TD children using frontal and central parietal connectivity differences. The selected channels classified a subject as an ASD or TD with 85.5% classification accuracy with selected features and classification algorithm.

### C. Feature Extraction Engine

The feature extraction requires the identified features of Kurtosis Value (KTV) and Katz Fractal Dimension (KFD) in the beta (12-30 Hz) frequency band. KTV provides information about the degree of concentration of the signal around the mean [14]. KFD provides information about the energy decay of a signal [15]. Equations (1)-(2) define KTV and KFD. The 10-bits digitized and pre-processed (P.Process) EEG data sampled at 250 Hz is forwarded to the feature extraction engine (FEE). The FEE passes the EEG signal from a bandpass filter and then calculates the KTV or KFD feature using the required hardware components.  $F_x$  represents the KTV or KFD feature for a single EEG channel.

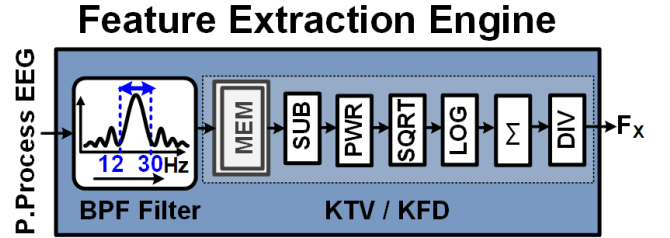


Fig. 3. Feature Extraction Engine highlighting a single channel

$$KTV = \sum_{i=0}^n \frac{(X_i - \bar{X})^4}{(N-1) * S^4} \quad (1)$$

$$KFD = \frac{\log(\sum ED(X_i, X_{i-1}))}{\log(\max(ecd(X_i, X_{i-1})))} \quad (2)$$

$X_i$ ,  $\bar{X}$ ,  $N$ , and  $S$  represent the time-series EEG sample, mean EEG value, total number of EEG samples, and standard deviation of the EEG data respectively. KTV (1) calculates the ratio of fourth power summation of differences of  $X_i$  and  $\bar{X}$  with a product of one less than the total number of samples ( $N-1$ ) and the fourth power of  $S$ . KFD (2) calculates the ratio between logarithms of summation of Euclidean differences (ED) between consecutive EEG samples ( $X_i$  and  $X_{i-1}$ ) and the maximum ED. The calculation of these features requires huge memory requirements ( $> 15$  MB) along with complex floating-point logarithm, power, and square root calculations (1)-(2). These calculations would make the ASD classification processor's hardware implementation unrealizable and impractical due to high power consumption ( $> 500$  mW) and huge silicon area requirements or FPGA resource constraints. Therefore, it is quite important to optimize these features to a hardware realizable approximation.

$$KTVI = K * SDI^4 \quad (3)$$

$$SDI = \frac{\max(X) - \min(X)}{4} \quad (4)$$

$$KFDI = \max(X_i - X_{i-1})^2 - \sum_{i=0}^n (X_i - X_{i-1})^2 \quad (5)$$

KTV (1) and KFD (2) were approximated to KTV indication (KTVI) and KFD indication (KFDI) respectively. KTVI (3) was calculated using the product of the fourth power of standard deviation indicator (SDI) and a constant parameter K. SDI is the approximated standard deviation using range rule [16]. The SDI simply requires the difference between the maximum and minimum samples in the EEG time series represented by  $\max(X)$  and  $\min(X)$ , respectively. Eq (4) shows the calculation of SDI where X represents the EEG data. The KFDI (5) calculates the difference between squares of the maximum difference and the total difference between  $X_i$  and  $X_{i-1}$ .  $X_i$  and  $X_{i-1}$  represent the current and previous EEG samples.

Fig. 4 shows the FEE to calculate the KTVI. The pre-processed EEG data in the beta frequency band (EEG  $\beta$  band) and the electrode/ channel number (ELT) were forwarded as input to the KTVI calculation unit. The EEG  $\beta$  band was calculated using a quantized FIR filter of 30th order as a half-precision (16'b) floating-point value.

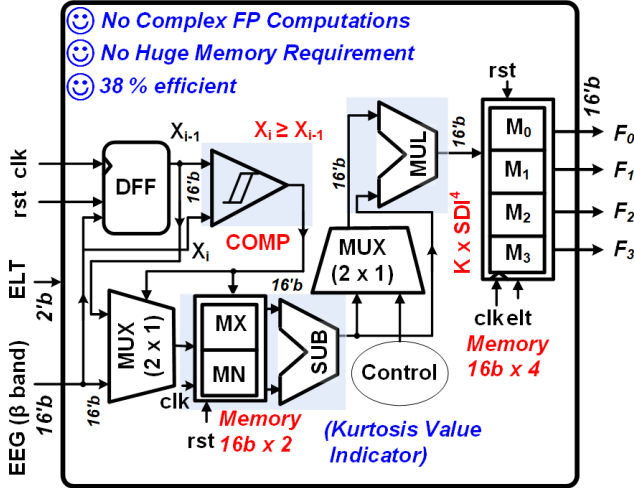


Fig. 4. FEE highlighting KTVI

The ELT represents the current electrode from the subset of four electrodes used for the classification. A comparator unit (COMP) compares the consecutive EEG samples and raises the output flag if the current EEG sample ( $X_i$ ) is higher than the previous sample ( $X_{i-1}$ ). The output flag of the comparator is used as a section input for a two-to-one multiplexer to update the contents of minimum (MN) and maximum (MX) values.  $X_i$  and  $X_{i-1}$  were sampled by a flip-flop (DFF) and forwarded as inputs to the multiplexer. A 32-bits memory unit block (16'b x 2) was used to store the MN and MX values. A FP subtractor (SUB) calculates the difference between MX and MN values. The fourth power SDI and KTVI were calculated using a single floating-point multiplication unit (MUL) controlled by the control unit. The KTVI of the selected channel was stored in a memory block (16'b x 4) using ELT.  $F_{0-3}$  represent the KTVI of the selected four channels. The proposed KTVI implementation does not require any complex FP calculations and huge memory requirements and was 38% efficient than conventional KTV implementation using (1).

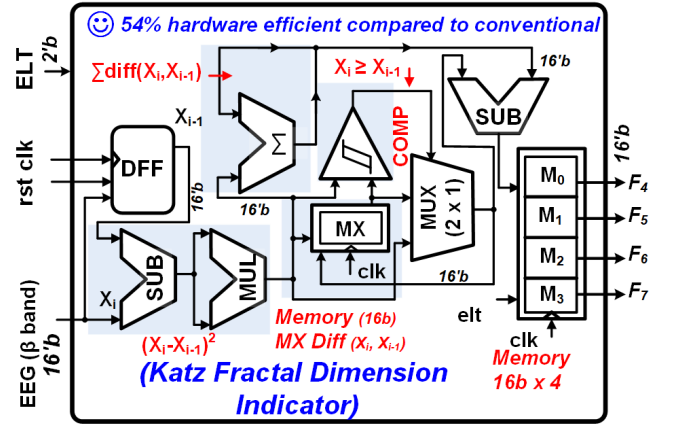


Fig. 5. Feature Extraction Engine highlighting KFDI

Fig. 5 shows the FEE to calculate the KFDI. The KFDI similarly requires the EEG  $\beta$  band and ELT.  $X_i$  and  $X_{i-1}$  were sampled using a DFF and the difference between  $X_i$  and  $X_{i-1}$  was calculated similarly to KTVI. A 16-bits floating-point summation unit ( $\Sigma$ ) was used to calculate the summation of differences (5) between  $X_i$  and  $X_{i-1}$ . The maximum difference (MX Diff ( $X_i, X_{i-1}$ )) was calculated using a floating-point comparator (COMP) controlled by a 2-1 multiplexer (MUX). MX was stored in a 16-bits memory block and updated using a MUX controlled by the COMP. A floating-point subtraction unit (SUB) and multiplication unit (MUL) were used to calculate the squared difference (5) between  $X_i$  and  $X_{i-1}$ . The KFDI of the selected channel was similarly stored in a memory block (16'b x 4) using ELT.  $F_4, F_5, F_6$  and  $F_8$  represent the KFDI of the selected four channels. The proposed KFDI implementation does not require any complex FP calculations and huge memory requirements and was 54% efficient than conventional (2) KFD implementation. The calculated features  $F_{0-7}$  were forwarded to the SNN classification unit after normalization as a feature vector.

#### D. Shallow Neural Network Classification Unit

A Shallow Neural Network (SNN) is a fully connected neural network without multiple hidden layers. The SNN classifies the output as ASD and TD by adjusting or optimizing the weights and biases during the learning process from the difference between the desired output and the actual output through backpropagation.

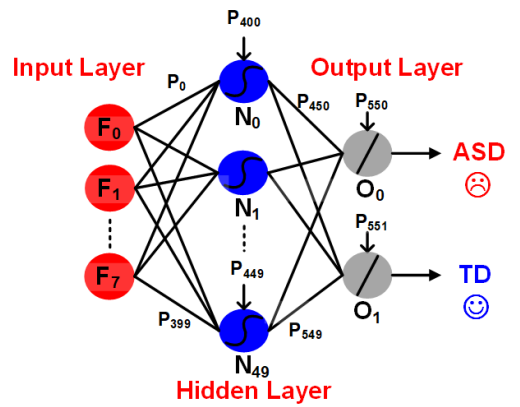


Fig. 6. SNN architecture

Fig. 6 shows the architecture of the SNN used for ASD classification. The SNN contains eight, fifty, and two nodes



in the input, hidden, and output layers respectively. The eight normalized features ( $F_0$ - $F_7$ ) are forwarded to the input layer.

$$N_{0-49} = \text{Sigmoid}\left(\sum_{a=399, b=7, c=449}^{a=459, b=49, c=551} P_a \cdot F_b + P_c\right) \quad (6)$$

$$O_{0-1} = \sum_{a=450, b=0, c=550}^{a=450, b=0, c=550} (P_a \cdot N_b + P_c) \quad (7)$$

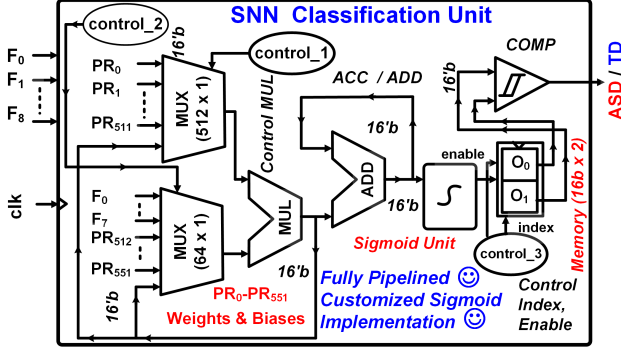


Fig. 7. SNN classification unit hardware architecture

The input layer (6) calculated the hidden layer values ( $N_{0-49}$ ) using multiplications and additions with the parameters ( $P_{0-449}$ ) and a sigmoid function. The output layer (7) values ( $O_{0-1}$ ) are calculated using  $N_{0-449}$  and output layer parameters ( $P_{450-551}$ ). The higher value of  $O_0$  or  $O_1$  classifies the patient as ASD or TD respectively. Equations (6)-(7) represent the mathematical operations required for SNN implementation.  $P_{0-399}$  and  $P_{400-449}$  are the weights and biases for the input layer respectively.  $P_{450-459}$  and  $P_{550-551}$  are the weights and biases for the output layer respectively.

Fig. 7 shows the hardware implementation of the SNN classification unit. The normalized features  $F_{0-7}$  are inputted to the classification unit, which uses a floating-point multiplier and adder to perform the addition or accumulation functions (6)-(7). Two multiplexers (512 x 1 and 64 x 1) are used to select multiplier's inputs to perform the multiplication, accumulation, or addition functions. Two finite state machine control units (control\_1 and control\_2) are used to provide the selection inputs of the multiplexers. A sigmoid unit is used to apply the sigmoid activation function [2]. A 32-bits memory block is used to store  $O_0$  and  $O_1$ . A finite state machine control unit (control\_3) is used to control the memory block using index and enable. The classification output (ASD/TD) is calculated after comparing  $O_0$  and  $O_1$  using a floating-point comparator.

### III. RESULTS AND DISCUSSION

The proposed ASD classification processor is implemented on Xilinx Atrix-7 FPGA. This work is the first hardware-based implementation of an ASD classification processor to the best of our knowledge verified on a dataset of ASD patients. The EEG dataset for ASD classification by [6] was used for this work. The overall power of 8.62  $\mu$ W is consumed while operating at 100 MHz clock. The processor classifies a patient as ASD or TD with 85.5% classification accuracy. The classification results were evaluated using a 5-fold cross-validation scheme.

Table 1. Comparison with the state-of-the-arts

	IEEE ISCAS'19	IEEE TCASII'15	IEEE Access'20	Nature SR'18	IEEE IRI'19	This Work
H/W	Yes (FPGA)	Yes (SoC)	Yes (FPGA)	No	No	Yes (FPGA)
Power	12.3 uW	13.6 uW	150 mW	NA	NA	8.62 uW
Electrodes Count	8 *	16 *	14 *	19	32	4
Accuracy	63 % *	100 % *	83.1 % *	95 %	95.5 %	85.5 %
LUTs	—	—	26229	NA	NA	18361
FFs	—	—	15180	NA	NA	10627
Application	Emotion	Epilepsy	Emotion	ASD	ASD	ASD

The comparison of the work with previous ASD classification processors [6], [9] is shown in Table 1. Since no other hardware-based ASD classification processor exists, the results are also compared with similar systems for other biomedical applications [8],[12],[13]. The classification performance of the work is quite good (85.5%) being the 1<sup>st</sup> hardware implementation and using the lowest number (4) of electrodes. Since the other hardware implementations target different biomedical applications, the classification accuracy does not represent a lateral comparison alone. The overall classification power, lookup tables and flip-flops count is significantly lesser than epilepsy or emotion classification.

### IV. CONCLUSION

Wearable ASD classification processors can be a major breakthrough in biomedical healthcare. They would assist ASD children and their caregivers in ASD diagnosis without any feeling of stigma. The implemented SNN classification processor utilizes the approximated and optimized implementations for hardware costly KTV and KFD features with 38% and 54% lesser hardware resources compared to conventional implementation. The high classification results and lower hardware resources are quite encouraging to develop a fully integrated SoC system for ASD classification after validation of the system after incorporating more ASD datasets.

### REFERENCES

- [1] "Autism Speaks" [Online]. Available: <https://www.autismspeaks.org/autism-statistics>
- [2] A.Aslam, T. Iqbal, M. Aftab, W. Saadeh and M.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)*, March. 2020.
- [3] A. McCrimmon and K. Rostad, "Test Review: Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part II): Toddler Module", *Journal of Psychoeducational Assessment*, vol. 32, no. 1, pp. 88-92, December. 2013.
- [4] E. Fuller and A. Kaiser, "The Effects of Early Intervention on Social Communication Outcomes for Children with Autism Spectrum Disorder: A Meta-analysis," *Journal of Autism and Developmental Disorders*, vol. 50, no. 1, pp. 1683-1700, May. 2020.
- [5] W. Saadeh, F. H. Khan and M. Altaf, "Design and Implementation of a Machine Learning Based EEG Processor for Accurate Estimation of Depth of Anesthesia," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 4, pp. 658-669, August. 2019.
- [6] 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)*, August. 2019, p. 151-158.

- [7] A. Aslam and M. 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, August. 2020.
- [8] A. Aslam, and M. Altaf, "An 8 Channel Patient-Specific Neuromorphic Processor for the Early Screening of Autistic Children through Emotion Detection," *IEEE International Symposium on Circuits and Systems (ISCAS)*, May. 2019.
- [9] W. Bosl, A. Tierney, H. Tager-Flusberg and C. Nelson, "EEG Analytics for Early Detection of Autism Spectrum Disorder: A data-driven approach", *Scientific Reports*, vol. 9, no. 6828, May. 2018.
- [10] M. Christ, N. Braun, J. Neuffer and A. Kempa-Liehr, "Time Series Feature Extraction on basis of Scalable Hypothesis tests" , *Neurocomputing*, vol. 307, pp. 72-77, September. 2018.
- [11] M. Olson, A. Wyner and R. Berk, "Modern neural networks generalize on small data sets" in *Advances in Neural Information Processing Systems*, Red Hook, NY, USA:Curran Associates, pp. 3619-3628, December. 2018.
- [12] M. Shoaran, C. Pollo, K. Schindler and A. Schmid, "A Fully Integrated IC With 0.85- $\mu$ W/Channel Consumption for Epileptic iEEG Detection," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 62, no. 2, pp. 114-118, February. 2015.
- [13] 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, July. 2020.
- [14] F. Al-Athari, "Confidence Interval for Locations of Non-kurtosis and Large Kurtosis Leptokurtic Symmetric Distributions", *Journal of Applied Sciences*, vol. 11, no. 3, pp. 528-534, 2011.
- [15] D. R. Jevtić and M. P. Paskaš, "Application of Katz algorithm for fractal dimension in analysis of room impulse response," in *2011 19th Telecommunications Forum (TELFOR) Proceedings of Papers*, November. 2011, pp. 1063-1066..
- [16] X. Wan, W. Wang, J. Liu and T. Tong, "Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range", *BMC Medical Research Methodology*, vol. 14, no. 1, December. 2014.