2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)
Amity University, Noida, India. Oct 13-14, 2022

2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) | 978-1-6654-7433-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICRITO56286.2022.9964717

Pre-processing of EEG signal using Independent Component Analysis

Bhawna kaliraman
ECE Deptt.
DPGITM
Gurugram, India
bhawna.kaliraman5@gmail.com

Sweety Nain
ECE Deptt
PM College of Engg.
Sonipat, India
sweetynain28@gmail.com

Rashmi Verma
CSE Deptt
DPGITM
Gurugram, India
itsrashmiverma@gmail.com

Monika Thakran
CSE Deptt
DPGITM
Gurugram, India
thakranmonika04@gmail.com

Yash Dhankhar
CSE Deptt
DPGITM
Gurugram, India
yashdhankhar92@gmail.com

Parli B Hari
CSE Deptt
DPGITM
Gurugram, India
parlibhari@gmail.com

***Abstract*- Biometric systems play vital role in everyday life. After signal acquisition, first step for any biometric system is to process the signal and extract valuable information from that signal. Pre-processing the EEG signal, which is essentially a collection of signal processing techniques before the primary EEG data analysis, is required to extract just brain activity from noisy EEG data. The selection of pre-processing techniques can influence the results of future EEG data analysis. Pre-processing of EEG signal includes various techniques like reduction of line noise, removal of redundant channels and reducing artefacts from EEG signal. EEG data is highly contaminated from various artefacts. This paper implements ICA as a pre-processing technique to remove eyeblink artefacts from signals which is then followed by Auto Regressive model features and results are obtained using Support Vector Machine classifier. Results shows that EEG signals performs better after applying ICA to raw data.**

Keywords- Electroencephalogram (EEG), Independent component analysis (ICA), Autoregressive (AR), Support Vector Machine (SVM)

I. INTRODUCTION

EEG signals are acquired in raw form and those signals incorporates noise. Firstly, noise is removed from them for further processing. So, after acquiring signals next step is to enhance the EEG signals to convert them into stage which is suitable for feature extraction. Pre-processing of EEG signals is also required to eliminate other artifacts from the raw signals. EEG signals have several frequency bands but for user identification we may have interest in certain frequency band only. So, filtering has been done to extract that particular frequency band. Different frequency bands are shown in table 1.

TABLE I. FREQUENCY BANDS FOR EEG SIGNALS

Rhythm	Frequency band
Delta (δ)	< 4 Hz
Theta (θ)	< 7 Hz
Alpha (α)	8-15 Hz
Beta (β)	16-31 Hz
Gamma (γ)	>32 Hz

A. EEG Artifacts

Artifacts are unwanted potentials that alter brain impulses. These are often derived from non-cerebral origins. These are derived from both physiological and non-physiological sources. Non-physiological sources are those that originate outside of the human body and may be readily removed by filtering, shielding, and other means. Physiological sources are those that arise from various body processes, like the impedance of electrodes, which can fluctuate according to skin responses such as perspiration, which causes distortions in the EEG data. The undesirable potential that affects brain signals is called artifacts.

- EOG (Electrooculogram) is a kind of physiological artifact which resulted because of blinking of eyes and it has high amplitude values in EEG data, these are also caused by movement of eyes which appears as a low frequency pattern.
- Another type of physiological artifact is EMG (Electromyogram), which is generated by movement of the head, torso, jaw, or tongue and produces severe disorders in EEG data.

To avoid the problem of physiological artefacts, ICA i.e., Independent Component Analysis can be used.

II. LITERATURE REVIEW

To extract particular set of frequencies different filters are also used in processing of EEG signal. Different pre-processing techniques commonly used by researchers are briefly explained in Table II. Table II contains various columns which contains information about authors of particular research which is followed by year of publication then number of subjects considered for particular research then number of EEG channels used to acquire EEG data from users then Frequency band used then next column shows the techniques used for pre-processing of EEG signal which is then followed by efficiency values obtained by using particular set of techniques.

TABLE II. SUMMARY OF VARIOUS PRE-PROCESSING METHODS USED IN LITERATURE FOR EEG SIGNALS

Author	Year	Subjects	EEG channels	Band	Pre-processing technique	Efficiency
Polous et al. [1]	2002	4 + 75 intruders	1 (O2)	All	Low pass filter (1-30) Hz	76% to 88%
Palanappan and Ravi [2]	2006	20	61	γ	PCA	96.5%
Riera et al. [3]	2008	51 + 36 intruders	2 (Fp1, Fp2)	All	2^{nd} order pass band filter (0.5-70) Hz, narrow notch filter (50Hz)	98.1%
Yang and Deravi [4]	2013	50	5	All	Noise elimination using wavelet and hybrid de-noising	95.5%
Fraschini et al. [5]	2015	109	64	All	BPF (0.5-50) Hz	99.5% (REO), 98.1% (REC)
Koike-Akino et al. [6]	2016	25	14	All	BSS-CCA	96.7%
Nishimoto et al. [7]	2020	20	64	All	BPF (0.5-40 Hz), CAR (Common Average Referencing)	50%
Leon et al. [8]	2020	20	16	Delta, theta, alpha, beta	BPF (0.3-30) Hz	91.67%
D. Carrion et al. [9]	2021	32	32	all	BPF (4-45Hz), BSS, DWT	95.54%
M. Wang et al. [10]	2022	109	64	All	FIR filter with hamming window, ICA and MARA	96.7%

III. METHODOLOGY

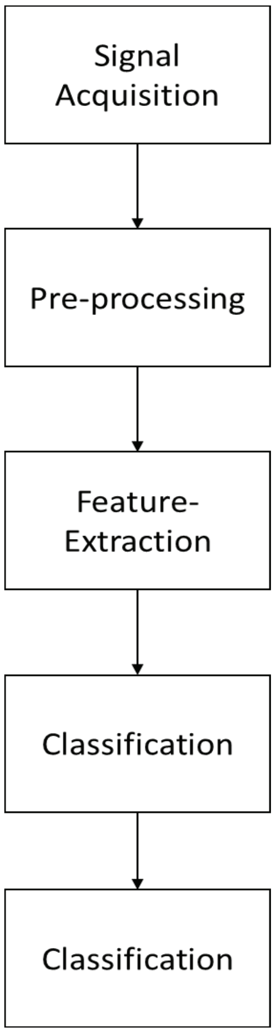

Fig. 1. Flowchart of methodology used

Figure 1 shows the flowchart of methodology used in present research. First step is the acquisition of EEG signals. After signal acquisition next step is pre-processing those acquired signals, so that they can be used for further processing. Then feature extraction is done in which relevant features are extracted from pre-processed EEG data. After extraction of features, these features are classified using suitable classifier which classifies between various users. This section contains experimental setup details, pre-processing technique used, feature extraction technique used and classifier used.

A. Experimental Setup And Simulation

To understand the EEG acquisition process, BIOPAC system has been used to acquire data. Also, due to lack of availability of unprocessed data, there is need to collect a raw dataset which can be used for pre-processing techniques. After studying various data acquisition techniques, we have acquired data using resting state with eyes open condition. The dataset is acquired from BIOPAC MP100 EEG data acquisition system. This system contains 32 channels to acquire data. Data is acquired using standard 10-20 electrode placement configuration. Ten healthy subjects were considered for data acquisition. Subjects are aged between (20-30 years). For acquisition of data, subjects have to sit and relax in calm environment on chair with their eyes open. The system setup is done under laboratory environment.

BIOPAC MP100 systems gives USB ready data acquisition and analysis. It can be used to record data from various channels with different samples rates. The MP100 includes a built-in microprocessor that handles data gathering and communication to the device.

B. Pre-Processing Using Independent Component Analysis (ICA)

ICA is a BSS ("Blind Source Separation") approach for identifying statistically independent, non-gaussian variables or components in multivariate data. This

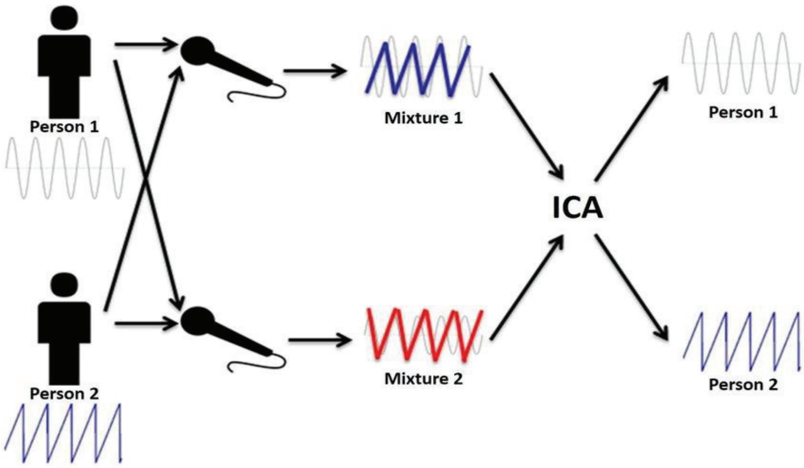

Fig. 2. ILLUSTRATION OF ICA

method's study goal is a non-gaussian signal. The origins of ICA theory may be traced back to the early 1990s, when French researchers Jutten and Hérault [11] introduced the notion of ICA for the first time.

“ICA can be explained with cocktail party problem where a mixture of sound (X1 and X2) is produced from a variety of sources (P1 and P2). The original sources (i.e., P1 and P2) can be retrieved back from mixer signal X1 or X2 with the application of ICA”[11] (Figure 2)

ICA have been applied on acquired EEG signals to reduce eyeblink artefacts.

C. *Feature Extraction*

Autoregressive Model: In AR model, previous values are used to predict the variable of interest. Auto regressive burgs model is used. It is calculated using Equation 1.

$$Y_t = a + \sum_{k=1}^{n} \varphi_k Y_{t-i} + \varepsilon_t \quad (1)$$

Above equation is known as AR(n) model. Where n denotes the order, a is constant, ε_t is white noise and φ_k is model coefficient. Number of values used to predict variable of interest are based on order of model.

D. *Classification*

Support vector machines (SVM): It is a supervised ML model. Its major goal is to find the hyperplane in N-D space (N represents no. of features); this hyperplane serves as a decision boundary. Support vector machines employ kernel functions to find support vector classifiers in higher dimensions consistently. To decrease computation, the support vector machine employs a kernel technique. The kernel trick refers to the method of computing a higher dimensional connection without actually changing the input to a higher dimension. For this project, a linear kernel was used.

IV. RESULTS AND DISCUSSION

In this section results are discussed which are obtained after applying pre-processing on raw EEG data. Auto regressive model with order three has been used for acquiring features and SVM is used for the classification. Fivefold cross validation has been used for classification. Data from each user is compared with data from other user. Classifier classifies between different users like biometric systems. Figure 3 shows the results before and after applying ICA to raw EEG data.

TABLE III. EFFICIENCY VALUES WITH ICA AND WITHOUT USING ICA

	Fold	Recall	Precision	Kappa	Accuracy
Without ICA	1	0.82156	0.82368	0.82469	0.82843
	2	0.81532	0.81236	0.81244	0.81341
	3	0.82231	0.82451	0.82624	0.82563
	4	0.83451	0.83653	0.83633	0.83726
	5	0.83564	0.84412	0.83313	0.84665
With ICA	1	0.84561	0.84566	0.84563	0.85136
	2	0.85142	0.85432	0.85124	0.85639
	3	0.86632	0.86415	0.86621	0.86763
	4	0.87332	0.87345	0.87631	0.87345
	5	0.87432	0.87542	0.87262	0.87431

Table III shows the efficiency values obtained after applying ICA and before applying ICA to EEG data. Various performance metrices such as Recall, Precision, Kappa and Accuracy have been used to validate the output. Further, five- fold cross validation has been used and results for each fold has been shown in table below. Results shows that efficiency values have been increased after applying ICA to EEG data. Figure 4 shows the chart representing various performance values of different performance metrices with different fold.

Results shows that model performs better with ICA as compared to without ICA because it removes redundant artifacts from signal.

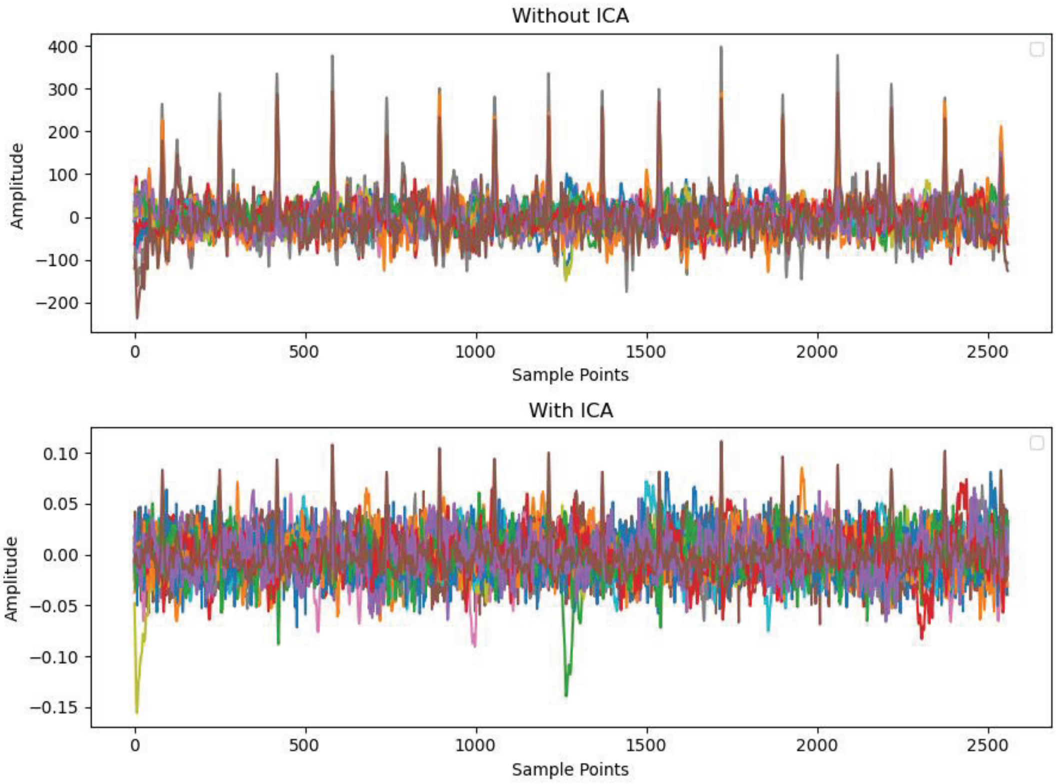

Fig. 3. EEG SIGNALS BEFORE AND AFTER APPLYING ICA

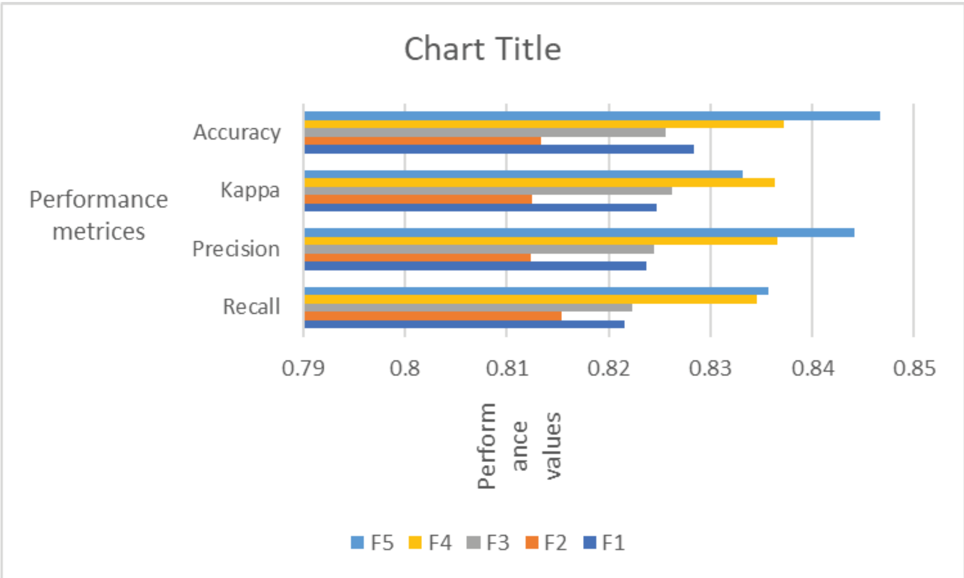

Fig. 4. PERFORMANCE CHART OF VARIOUS PERFORMANCE METRICES.

V. Conclusion And Future Work

In this work, preliminary investigation on the effectiveness of the pre - processing stage for based EEG biometrics have been presented. Independent component analysis-based strategies for artifact removal have been implemented on database acquired using MP100 system. Data has been collected form 10 users in EO state. Features have been extracted using AR model from pre-processed EEG signals and SVM with linear has been used for classification purpose. Experimental results as show high recognition rates achieved using ICA as pre-processing technique as compared to without ICA. In future, various techniques for extraction of features and selection of extracted features can be implemented to obtain higher efficiency. Also, number of users can be increased for data acquisition.

References

[1] M. Poulos, M. Rangoussi, N. Alexandris, and a Evangelou, "Person identification from the EEG using nonlinear signal classification.," *Methods of information in medicine*, vol. 41, no. 1, pp. 64–75, 2002.

[2] R. Palaniappan and K. V. R. Ravi, "Improving Visual Evoked Potential Feature Classification for Person Recognition Using PCA and Normalization," *Pattern Recognition Letters*, vol. 27, no. 7, pp. 726–733, 2006, doi: 10.1016/j.patrec.2005.10.020.

[3] A. Riera, A. Soria-Frisch, M. Caparrini, C. Grau, and G. Ruffini, "Unobtrusive Biometric System Based on Electroencephalogram Analysis," *Eurasip Journal on Advances in Signal Processing*, vol. 2008, no. 18, 2008, doi: 10.1155/2008/143728.

[4] S. Yang and F. Deravi, "Wavelet-based EEG preprocessing for biometric applications," *Proceedings - 2013 4th International Conference on Emerging Security Technologies, EST 2013*, pp. 43–46, 2013, doi: 10.1109/EST.2013.14.

[5] M. Fraschini, A. Hillebrand, M. Demuru, L. Didaci, and G. L. Marcialis, "An EEG-based biometric system using eigenvector centrality in resting state brain networks," *IEEE Signal Processing Letters*, vol. 22, no. 6, pp. 666–670, 2015, doi: 10.1109/LSP.2014.2367091.

[6] T. Koike-akino *et al.*, "High-Accuracy User Identification Using EEG Biometrics," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 854–858.

[7] T. Nishimoto, H. Higashi, H. Morioka, and S. Ishii, "EEG-Based Personal Identification Method Using Unsupervised Feature Extraction and Its Robustness Against Intra-Subject Variability," *Journal of Neural Engineering*, vol. 17, no. 2, 2020, doi: 10.1088/1741-2552/ab6d89.

[8] J. León *et al.*, "Deep learning for EEG-based motor imagery classification: Accuracy-cost trade-off," *PLoS ONE*, vol. 15, no. 6, pp. 1–30, 2020, doi: 10.1371/journal.pone.0234178.

[9] D. Carrión-Ojeda, R. Fonseca-Delgado, and I. Pineda, "Analysis of factors that influence the performance of biometric systems based on EEG signals," *Expert Systems with Applications*, vol. 165, no. September 2020, p. 113967, 2021, doi: 10.1016/j.eswa.2020.113967.

[10] M. Wang, S. Wang, J. Hu, and S. Member, "Cancellable Template Design for Privacy-Preserving EEG Biometric Authentication Systems," pp. 1–14.

[11] C. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, no. 1, pp. 1–10, 1991, doi: https://doi.org/10.1016/0165-1684(91)90079-X.