ijamrsd.com JAMRSD ISSN: 2583-6404
Journal of Advanced Multidisciplinary Research Studies and Development
Peer Reviewed Referred Journal Bi-monthly Open Access

ANALYSIS OF DISCRIMINATIVE EEG SIGNAL FEATURES FOR RECOGNITION OF CONFUSING MARATHI DIGITS

Rahul D Chaudhari[1], Rakesh S Deore[2]
[1,2] S.S.V.P.S.L.K. Dr. P.R.Ghogrey Science College, Department of Computer Science Dhule- 424001, Maharashtra, India.
Email: merahulchaudhari@gmail.com[1], rakeshsdeore@gmail.com[2]

Abstract
Electroencephalography (EEG)-based brain–computer interface (BCI) systems enable direct communication between the human brain and external devices by interpreting neural signals. However, accurate classification of visually similar Marathi digits remains challenging due to overlapping neural activation patterns. In particular, Marathi digits २, ३, and ६ exhibit significant confusion during EEG-based recognition due to their similar visual complexity and corresponding neural responses. This study presents a comprehensive analysis of discriminative EEG features for recognizing these confusing digits. This works includes various preprocessing techniques for EEG signal filtering and artifact removal, feature extractions, dimensionality reduction and classification. Experimental results demonstrate that hybrid features combined with ICA and LDA significantly improve class separability and classification accuracy. The findings provide insights into neural feature discriminability and contribute to improving EEG-based Marathi digit recognition systems.
Keywords: EEG, Brain–Computer Interface (BCI), Marathi Digit Recognition, Feature Extraction, Classification, ICA, LDA.

► *Corresponding Author: Rahul D Chaudhari*

I. Introduction

This paper presents some of the most representative studies and applied protocols in the field of brain–computer interfaces, used for the Marathi number recognition using EEG signals. The aim is accurate classification of confusion Marathi digits which is challenging due to overlapping neural activation pattern.

One instrument that makes it possible to communicate with people without using voice or physical contact is a human brain–computer interface system [1]. It functions by using signal processing techniques to translate brain activity into commands for a computer system. EEG records the electrical activity produced by the different brain regions by tracking changes in the voltage of the ionic influx into the neurons. The user's ideas are then shown in real time on a screen by the computer system. It avoids the usual sensory or motor channels and provides a direct line of communication between an external device and the brain of a person with cognitive, social, or sensorimotor deficits [2]. Every group that gets a BCI has a neurological disorder or impairment [3, 4].

The neurophysiologic measurement of the brain's electrical activity is called electroencephalography (EEG) [5]. This signal is often expressed in terms of amplitude, frequency, and phase and is a function of time [6]. Electrical impulses are used by the neurons to communicate, and they produce a bio-electromagnetic field that travels through the scalp, skull, and brain tissues [7]. The detectors are positioned on the scalp to simultaneously monitor signals that characterize brain activity from several places [8].

II. Related Work

Using machine learning approaches, EEG-based categorization has been extensively investigated for the recognition of cognitive states, visual stimuli, numbers, and motor imagery tasks.

A visual number identification system based on EEG signals was proposed by D. Nath, M. B. Uddin et al. (2015). Different visual stimuli produce unique brain patterns that may be identified from EEG recordings.Using frequency-domain and time-frequency domain analysis to extract EEG characteristics, the study discovered that the beta rhythm outperformed other EEG rhythms in number recognition.The findings also shown that Power Spectral Density (PSD) analysis of the beta rhythm yields more unique patterns than FFT, which makes it more appropriate for precise number identification and helpful in applications like robotics and neurological condition investigation [9].

In their evaluation of the best feature extraction technique for distinguishing between decimal numbers, D. Nath and M. Ahmad (2015) showed that the best classifier had a better recognition rate. It was discovered that PSD analysis of beta rhythm is more appropriate than FFT for more accurate number identification, and frequency domain analysis offers a greater recognition rate than time domain analysis. The sigmoid kernel outperforms other classifiers in terms of recognition rate and achieves 93.75% reliability for PSD analysis, surpassing DWT analysis. PSD analysis can clearly distinguish and offer more distinctive patterns for a given number[10].

The issue of handwritten digit recognition, which is crucial in applications like bank check processing and car number plate identification, was examined by Alomari, Saleh & Putra, Sumari et al. (2009). The goal of the project was to employ image processing and pattern recognition techniques to enable computers to detect Arabic numerals written by hand. Their technique effectively enabled the computer to recognize handwritten Arabic digits supplied by users by utilizing a segmentation method and neural network methodology[11].

The recognition of Arabic handwritten numerals (Old-Indian numbers) employed in text recognition applications was investigated by Marwan A. Abu-Zanona, Bassam et al. (2012). Despite being simpler than Arabic symbols, these numerals' curved curves make it challenging for intelligent systems to recognize them. To extract information and identify numbers, the study employed a neural network model and a morphological segmentation technique. The system was tested using a MATLAB-based application that achieved around 98% reliability on the test dataset[12].

The goal of S.L. Mhetre and Prof. M. M. Patil's (2013) study is to identify Devanagari numerals by grid-based and artificial neural network techniques. The ANN's performance is dependent on how much it has been trained; if the number of samples is increased, it may reach up to 96%. Additionally, the techniques may be used to offline signature verification and printed or handwritten Devanagari character identification[13].

Using EEG signals, Rao Shashibala and Gawali Bharti et al. (2012) investigated number recognition. For the experiment, a GUI random number generator based on VB7 was created to show the digits 0–9 as visual stimuli. Six right-handed male individuals between the ages of 20 and 25 provided EEG data for the training dataset, while fresh recordings from the same subjects served as the testing dataset. The system achieved an overall identification accuracy of 68.33% by focusing on the beta signal and classifying using Linear Discriminant Analysis (LDA)[14].

The work of Paul Jaya and Anasua Sarkar (2018) demonstrates the use of CNN for handwritten Bangla number recognition. With an accuracy of 99.2%, deep learning approaches may get a high level of accuracy. Although the suggested approach did not test on many datasets of handwritten Bangla characters, it might produce better results with more data and a larger CNN network design[15].

Some of the peoples were worked on character recognition using EEG Signal Processing are as follows.

A innovative method of communicating with the outside world using human brain activity is the letter recognition system (S. Wahed, M. Islam et al., 2015). It is based on temporal and spatial analysis to identify key characteristics in the unprocessed EEG data, which are further examined using wavelet and FFT. For statistical and wavelet analysis, the overall recognition rates are 80% and 85.6%, respectively. It is demonstrated that the suggested approach can effectively and consistently recognize the English alphabet[16].

According to S. Wahed, M. Rana, et al. (2016), the best wavelet for letter recognition was found using brain signal analysis. SVM was utilized to categorize the ten characteristics that were taken out of the sub-bands. With an overall accuracy rating of 90%, Daubechies of order 8 was determined to be the optimal wavelet family. At 80%, delta rhythm had the highest recognition rate. The most effective analytic technique for precisely classifying EEG data for letter recognition is Daubechies 8[17].

III. Methodology

In this research we collect EEG signal while participate seeing random marathi digits (२, ३, and ६) for every 2 seconds until 2 minutes. We apply various EEG signal preprocessing techniques on the raw data after that we apply feature extraction techniques for feature extraction. We use dimensionality reduction for exacted features and use classification for accurately recognized confusion marathi digits.

A. EEG Data Acquisition

There is no dataset available for Marathi digit recognition using EEG signal processing. Therefore, in the following Figure 1, we are set up an EEG data collection model in which EEG electrodes placed by using a 10 -20 international system on the brain, and the subject sees Marathi numbers ० - ९ randomly generated by Python Tkinter GUI program 2 minutes time frame on a computer screen.

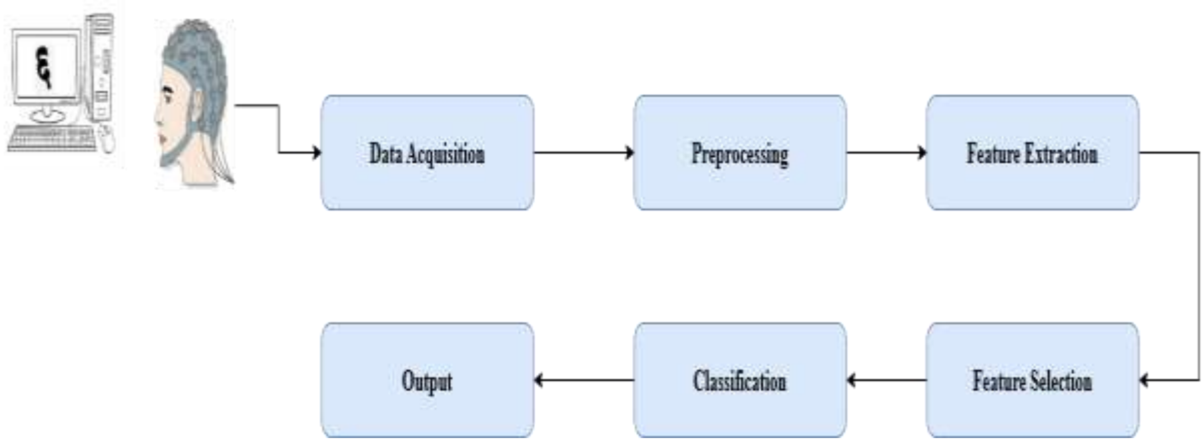


Figure 1:- EEG Data Collection System

Using this model, we collect EEG signals while the subject sees random Marathi digits on a computer screen for a 2-minute time frame. We use one of the greatest 32-channel EEG machines, the RMS Super Spec. The purpose of the RMS Super Spec 32-channel EEG Machine is to work with a desktop computer.

B. Preprocessing

EEG raw data has noises and artifacts there is need to remove that noises and artifacts formrecorded raw data for this we use band pass filter, notch filter and Independent Component Analysis (ICA)

1. Bandpass Filter

The raw EEG data were subjected to a bandpass filter in order to remove high-frequency noise and low-frequency drift while preserving pertinent brain frequency components. Brain rhythms in EEG signal processing are usually found in certain frequency ranges, such as beta (13–30 Hz), alpha (8–13 Hz), theta (4–8 Hz), and delta (0.5–4 Hz). In order to improve the quality of the EEG signals before feature extraction and classification, a bandpass filter with a frequency range of 0.5–45 Hz was used to eliminate undesirable artifacts and noise while preserving relevant brain information.

$$H(f) = \begin{cases} 1, & f_L \leq |f| \leq f_H \\ 0, & \text{otherwise} \end{cases}$$

Where:

- f_L= lower cutoff frequency (e.g., 0.5 Hz)
- f_H= upper cutoff frequency (e.g., 45 Hz)

2. Notch Filter

In EEG preprocessing, a notch filter is used to eliminate a certain frequency range, often the 50 Hz power-line interference that is present in nations like Europe and India (or 60 Hz in the USA). Electrical noise from power lines and lab equipment that coincides with the EEG frequency range can be captured by EEG electrodes acting as antennas. The notch filter eliminates the 50 Hz noise while maintaining the crucial EEG brain wave bands since this interference lowers signal quality and classification accuracy.

The filter designed by iirnotch (w0, Q) has the transfer function:

$$H(z) = \frac{1 - 2\cos(\omega_0)z^{-1} + z^{-2}}{1 - 2r\cos(\omega_0)z^{-1} + r^2 z^{-2}}$$

Where:

$$\omega_0 = \frac{2\pi f_0}{f_s}$$
$$r = 1 - \frac{\pi}{Q}$$

Sampling frequency: $f_s = 256$ Hz

Notch frequency (India mains): $f_0 = 50$ Hz

Quality factor: $Q = 30$

Normalized frequency: $w_0 = \frac{f_0}{f_s/2}$

Because the bandwidth becomes narrowing and the filter's transient response lifetime grows as the radius of the pole of the notch IIR filter increases [18]. Here, we can see that a reduction in r improves the filter's transient responsiveness, or greater transient suppression. We can change the pole radius over time in order to shorten the IIR notch filter's transient response time [19].

3. Independent Component Analysis (ICA)

After the signals are linearly mixed with an unknown medium and recorded at N sensors, the independent source signals may be retrieved based on the blind source separation. Using contrast functions approximated by the Edgeworth expansion of the Kullback-Leibler divergence, the idea of independent component analysis (ICA) was defined as maximizing the degree of statistical independence across outputs [20]. ICA was used to eliminate artifacts from recorded EEG data, such as muscle noise and eye blinks.

EEG Mixing Model (Core Equation)

Let:

$X(t)$= observed EEG signals (channels × time)

$S(t)$= unknown source signals (brain + artifacts)

A= unknown mixing matrix

$$\boxed{X(t) = A\,S(t)}$$

Each EEG channel is a linear mixture of independent sources.

Goal of ICA

Find an unmixing matrix W such that:

$$\boxed{S(t) = W\,X(t)}$$

Where the components of $S(t)$ are statistically independent. [21]

C. Windowing

For effective analysis, windowing is used to split the continuous EEG signal into segments of a set length. This procedure makes it possible to extract significant information from every EEG data segment. Additionally, it boosts the quantity of training data, which enhances the classification model's performance.

D. Feature Extraction

To get detailed information about the brain signals, hybrid features were generated from each segmented EEG window. Time-domain, frequency-domain, and wavelet-based characteristics were all extracted. The statistical properties of the EEG signals were represented by time-domain metrics such mean, standard deviation, variance, and root mean square (RMS). By calculating the band power in several EEG frequency bands, such as delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz), which indicate various patterns of brain activity, frequency-domain properties were produced. The time-frequency properties of the EEG data were also captured by extracting wavelet-based features, specifically wavelet energy and wavelet entropy. Compared to single-domain features, these hybrid features offer a more complete picture of brain activity. The total number of features retrieved from each window is 160 since the EEG dataset has 16 channels and 10 features were extracted from each channel. Subsequent processing and categorization were done using these characteristics. The feature values were normalized using standard scaling.

E. Dimensionality Reduction

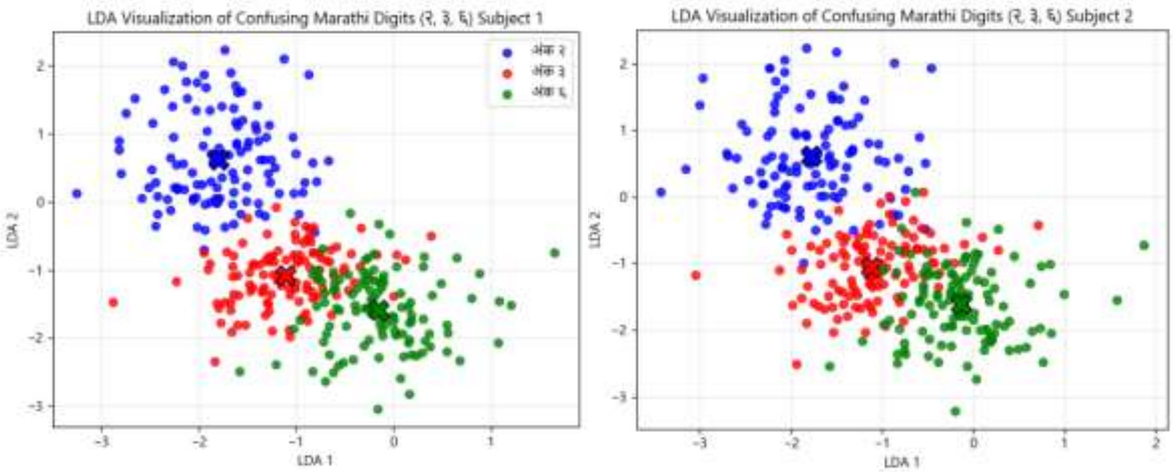
To simplify the collected feature set while maintaining the most discriminative information, dimensionality reduction was carried out using Linear Discriminant Analysis (LDA). Dimensionality reduction was required to reduce duplication and increase computing performance since the hybrid feature extraction approach generated a high-dimensional feature vector. By increasing the distance between distinct classes and decreasing the variation within the same class, LDA projects the original feature space onto a lower-dimensional space. This change improves the classifier's performance and increases the features' capacity for discrimination. In order to generate a compact feature representation that enables more accurate classification of Marathi digits, LDA was performed in this work following feature scaling.

F. Classification

Based on the acquired EEG data, a Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel was employed as the classifier to differentiate between confusing Marathi digits. SVM is a supervised machine learning technique that finds a hyperplane, or ideal decision boundary, to divide several classes in the feature space. Because brain signals are complex and extremely nonlinear, the classifier can describe nonlinear correlations between the EEG data and the digit classes thanks to the RBF kernel. Because SVM can efficiently handle high-dimensional feature spaces and limited training datasets, it is especially well-suited for EEG classification. In order to correctly understand the discriminative patterns linked to Marathi digit confusion, the SVM classifier in this work was trained using the reduced feature set that was created after using Linear Discriminant Analysis (LDA).

IV. Experimental Results

The outcomes of the experiment show how well the suggested EEG-based Marathi digit recognition system is. When it came to differentiating between confused Marathi numerals, hybrid feature extraction that included time-domain, frequency-domain, and wavelet-based features demonstrated high discriminative capabilities. Wavelet entropy and alpha band power were shown to offer the best separation between digit classes among the retrieved characteristics. By utilizing the LDA and SVM architecture, the suggested system was able to obtain excellent classification accuracy. However, because the brain activity patterns of the numerals २ and ३ are similar, there was some misunderstanding between them. Similar to this, the numbers ३ and ६ showed some overlap in the feature space, but the hybrid features were generally successful in differentiating between them. The difficulties with EEG-based visual digit identification were highlighted by confusion matrix analysis, which showed that most misclassifications were between visually identical digits. Additionally, LDA visualization showed that each Marathi digit corresponded to a discrete cluster with no overlap across perplexing digits, proving that the suggested preprocessing and feature extraction



V. Conclusion

This study demonstrated an EEG-based method for identifying Marathi numbers, with a special emphasis on digits that are difficult to understand. २, ३ and ६. Preprocessing, hybrid feature extraction, dimensionality reduction using LDA, and classification with SVM were all part of the suggested workflow. Despite little misunderstanding between similar digits, the experimental results demonstrated that the derived characteristics successfully differentiated the digits, attaining excellent classification performance. The work highlights the possibility of EEG-based methods for brain-computer interface applications including Marathi digit identification.

References

1. Van Kokswijk, J.; Van Hulle, M. Self-adaptive BCI as service-oriented information system for patients with communication disabilities. In Proceedings of the 4th International Conference on New Trends in Information Science and Service Science, Gyeongju, South Korea, 11–13 May 2010; pp. 264–269

2. Mattout, J. Brain-Computer Interfaces: A Neuroscience Paradigm of Social Interaction? A Matter of Perspective. Front. Hum. Neurosci. 2012, 6, 114.

3. Wolpaw, J.R.; Birbaumer, N.; McFarland, D.J.; Pfurtscheller, G.; Vaughan, T.M. Brain–computer interfaces for communication and control. Clin. Neurophysiol. 2002, 113, 767–791.

4. Vuckovic, A.; Pineda, J.A.; LaMarca, K.; Gupta, D.; Guger, C. Interaction of BCI with the underlying neurological conditions in patients: Pros and cons. Front. Neuroeng. 2014, 7, 42.

5. Al-Kadi MI, Reaz MB, Ali MA. Evolution of electroencephalogram signal analysis techniques during anesthesia. Sensors (Basel). 2013 May 17;13(5):6605-35. doi: 10.3390/s130506605. PMID: 23686141; PMCID: PMC3690072.

6. T. Hinterberger, N. Weiskopf, R. Veit, B. Wilhelm, E. Betta and N. Birbaumer, "An EEG-driven brain-computer interface combined with functional magnetic resonance imaging (fMRI)," in *IEEE Transactionson Biomedical Engineering*, vol. 51, no. 6, pp. 971-974, June 2004, doi: 10.1109/TBME.2004.827069.

7. Jarchi, Delaram &Boostani, Reza & Taheri, M. & Sanei, Saeid. (2009). Seizure source localization using a hybrid second order blind identification and extended rival penalized competitive learning algorithm. Biomed. Signal Process. Control.4.10.1016/j.bspc.2009.01.004.
8. Kamei S, Morita A, Serizawa K, Mizutani T, Hirayanagi K. Quantitative EEG analysis of executive dysfunction in Parkinson disease. J Clin Neurophysiol. 2010 Jun;27(3):193-7.doi: 1097/WNP.0b013e3181dd4fdb. PMID: 20461018.
9. D. Nath, M. B. Uddin, M. M. Rana, P. C. Biswas, S. Wahed and M. Ahmad, "Number recognition using salient features of beta rhythmic EEG signal," *2015 International Conference on Electrical Engineering and Information Communication Technology(ICEEICT)*, Savar, Bangladesh, 2015, pp. 1-6, doi: 10.1109/ICEEICT.2015.7307364.
10. D. Nath and M. Ahmad, "Toward number recognition system: A nonstationary signal analyzing approach through SVM algorithm," 2015 2nd International Conference on Electrical Information and Communication Technologies (EICT), Khulna, Bangladesh, 2015.
11. Alomari, Saleh & Putra, Sumari & Al-Taweel, Sadik & Husain, Anas. (2009). Digital Recognition usingNeural Network. Journal of Computer Science. 5.10.3844/jcssp.2009.427.434.
12. Marwan A. Abu-Zanona, Bassam M. El-Zaghmouri ,Current Arabic (Hindi) Hand Written Numbers Segmentaion and Recognition Advance ImageProcessing and Neural Network, Journal of EmergingTrends in Computing and Information Sciences, VOL. 3, NO. 6, June 2012, ISSN 2079-8407.
13. S.L.Mhetre1 ,Prof.M.M.Patil, Comparative study of two methods for Handwritten Devanagari Numeral Recognition, IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661. Dec.2013)
14. Rao Shashibala ,GawaliBharti , Rokade Pramod &Deore Rakesh (2012). Number recognition system using electroencephalogram (EEG) signals. Advances in Computational Research ISSN: 0975-3273,2012.
15. Paul, Jaya, and Anasua Sarkar. "Handwritten Bangla numeral recognition using convolutional neural networks." 2018 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech). IEEE, 2018.
16. S. Wahed, M. Islam, P. C. Biswas, M. M. Rana, D. Nath and M. Ahmad, "An approach for letter recognition system modeling based on prominent features of EEG," 2015 International Conference on Electrical & Electronic Engineering (ICEEE), Rajshahi, Bangladesh, 2015, pp. 5-8, doi: 10.1109/CEEE.2015.7428293.
17. S. Wahed, M. Rana, S. M. K. Hasan and M. Ahmad, "Toward letter recognition system: Determination of best wavelet and best rhythm using EEG," 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Dhaka, Bangladesh, 2016, pp. 1-4, doi: 10.1109/CEEICT.2016.7873099.
18. A. D. Ramesh Rajagopalan, "A Pole Radius Varying Notch Filter with Transient Suppression for Electrocardiogram," Int. J. Electron. Commun. Eng., vol. 8, no. 3, pp. 134–138, 2014
19. M. A. Oktar, "Offline And Real Time Noise Reduction In Speech Signals Using The Discrete Wavelet Packet Decomposition Graduate Faculty of Environment and Technology University of the West of England in partial fulfilment of the requirements for the degree of Master of ," no. October, 2017.
20. Comon P., Independent component analysis,A new concept? Signal Processing, l994,36: 287-3 14
21. Anem¨ uller J, Sejnowski T, Makeig S. Complex independent component analysis of frequency—domain electroencephalographic data. Neural Networks 2003;16:1311–1323.