IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 11 July 2022, accepted 31 August 2022, date of publication 12 September 2022, date of current version 19 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3205734

RESEARCH ARTICLE

Statistical Evaluation of Factors Influencing Inter-Session and Inter-Subject Variability in EEG-Based Brain Computer Interface

RITO CLIFFORD MASWANGANYI[1], CHUNLING TU[1], (Member, IEEE), PIUS ADEWALE OWOLAWI[1], AND SHENGZHI DU[2]

[1]Department of Computer Systems Engineering, Tshwane University of Technology, Pretoria 0183, South Africa
[2]Department of Electrical Engineering, Tshwane University of Technology, Pretoria 0183, South Africa

Corresponding author: Chunling Tu (duc@tut.ac.za)

ABSTRACT A cognitive alteration in the form of diverse mental states has a significant impact on the performance of electroencephalography (EEG) based brain computer interface (BCI). Such alterations include a change in concentration levels commonly recognized as being indicated by the alpha rhythm, drowsiness or mental fatigue which occurs during EEG signal acquisition. Change in mental state give rise to a challenge of variability in EEG characteristics across sessions and subjects. Consequently, this variability constitutes to low intention detection rate (IDR) that renders BCI performance unreliable. This study investigates the impact of multiple factors that lead to the poor performance of the EEG-BCI. Five factors 1) concentration level; 2) selection of independent components(IC); 3) inter-session variability; 4) inter-subject variability; and 5) classification methods on the IDR in EEG based BCI. The alpha rhythm, as the indicator of concentration level, is validated, and the relationship between the alpha rhythm and the IDR is studied among sessions. In addition, ICs are examined to determine their effects on the IDR across sessions. The possibility of two sessions to contain similar EEG characteristics is also examined, where both sessions are acquired from the same subject in different days. Moreover, the possibility of two different subjects to containing similar EEG characteristics is examined. Furthermore, to conquer the challenge of variability in EEG dynamics a feature transfer learning (TL) approach is proposed in this study. Furthermore, three classification methods (TL, K-NN and NB) are examined and compared to determine whether multi-source neural information can improve the classification accuracy of individual sessions or subjects. Three EEG datasets acquired using different paradigms are used for experiments. The datasets include steady state motion visual evoked potential (SSMVEP), motor imagery (MI) and BCI competition IV-a dataset. Experimental results have shown that selection of independent components has an effect on the IDR. In this case IC-2 and IC-11 achieved a lowest and highest accuracies of 51% and 100% for SSMVEP datasets, while IC-9 and the double-component (IC-2 and IC-13) achieved a lowest and highest accuracies of 40% and 69% for MI datasets respectively. The second experiment demonstrated that higher alpha rhythm, depicted by a lower IDR corresponds to a lower concentration level. While a lower alpha rhythm depicted by a higher IDR corresponds to a higher concentration level. Moreover, variability within sessions can significantly deteriorate intention detection rate across sessions. As such a decline in accuracy from 82% to 61%, and from 56% to 44% was observed across both SSMVEP and MI sessions during inter-session experiment respectively. Integration of samples from different sessions but same subject resulted in a highest accuracy of 65%, 59% and 40% for SSMVEP, MI and BCI competition dataset. Integration of samples from different subjects resulted in a highest accuracy of 65%, 44% and 48% for SSMVEP, BCI competition and MI datasets. When three classifiers are evaluated and compared to determine whether multi-source neural information can improve the classification accuracy of individual sessions and subjects or domains, both K-NN and NB

The associate editor coordinating the review of this manuscript and approving it for publication was Venkata Rajesh Pamula.

achieved highest accuracies of 59% and 52% respectively, while TL showed a significant increase with an accuracy of 98% achieved using SSMVEP sessions. In a similarly manner both K-NN and NB achieved highest accuracies of 49% and 42% respectively using SSMVEP subjects, while TL showed a significant increase with an accuracy of 64% achieved. Furthermore, when 9 MI subjects acquired from BCI competition dataset were used, both K-NN and NB achieved highest accuracies of 68% and 65% respectively, while a significant increase in accuracy was observed when TL is used with accuracy of 99% achieved. In conclusion, the change of alpha rhythm magnitude among sessions significantly affect the IDR across sessions. While component selection across sessions has significant effects due to non-linear and non-stationary nature of EEG signals. Moreover, merging of ICs from different sessions, and inter-subject factor introduce challenges of overfitting resulting in low IDR. The classification methods are also found critical, because some advanced classification methods can improve the classification accuracy.

INDEX TERMS Brain computer interface (BCI), alpha rhythm, motor imagery (MI), steady state motion visual evoked potential (SSMVEP), inter-session variability, inter-subject variability, transfer learning (TL).

I. INTRODUCTION

Brain-computer interface (BCI) is an interaction between the brain and external electronic devices using neural activities of the brain. The electroencephalogram (EEG) signals are commonly converted into control commands to control external devices. BCI makes it possible for people suffering from severe motor disability to communicate with the surrounding environment without the use of muscles, whereby a direct communication link between the brain and external devices is established to facilitate the communication [1]. In recent years several signal processing techniques and different experimentation paradigms (such as MI, SSVEP and Hybrid) have been introduced to enhance the performance of BCI systems [2], [3]. As such different signal filters and decomposition techniques have been used to eliminate the impact of both physiological and non-physiological artifacts on intention detection rate [4], [5], [6], while a wide variety of feature extraction and selection techniques have been utilized to determine relevant signal features [7], [8], [9], and enhance the performance of machine learning algorithms [10]. However, variabilities in EEG characteristics resulting from cognitive alterations still remains most significant challenge in BCI [11], [12], [13]. [14] illustrated through implementation of an MI based BCI system the impact and evidence of variabilities across different sessions, and also evaluated whether there is evidence of similarities between EEG dynamics of different subject [14]. In this case BCI competition IV-a dataset was utilized to implement two experiments (inter-session and inter-subject). These experiments were implement under three condition namely case 1 (22 channels), case 2 (9 channels from both sensorimotor and parietal area), and case 3 (9 channels from sensorimotor area only) [14], [15]. A band-pass filter was firstly applied on the dataset, while three spatial filters (common spatial pattern, regularized composite spatial covariance and Joint approximate diagonalization) were evaluated on filtered dataset. As such wavelet decomposition techniques was used to extract features (sub-band energy and entropy), while a Two-Layer Feed-Forward Neural Network was used to classify features. Consequently, a highest classification accuracy of 58% as achieved for inter-subject experiment using a pairwise performance associativity technique, while there was a 31% decrease in classification accuracy across sessions during inter-session experiment [14].

Reference [16] evaluated the inter-subject variability of BCI performance between paradigms and sessions, whereby three experimentation paradigms (MI, ERP and SSVEP) were considered to determine performance variations across sessions. Consequently, a highest average classification accuracy of 72.2%, 96.6% and 95.5% for MI, ERP and SSVEP was achieved respectively [16]. In this case for MI paradigm EEG signals were filtered using a 5th order butterworth band-pass filter (between 8-30 Hz), from which log-variance features were extracted using common spatial pattern (CSP), and classified using linear discrimination analysis (LDA) classifier to predict two MI classes [17]. For ERP paradigm EEG signals were filtered using a 5th order butterworth band-pass filter (between 0.5-40 Hz) [18], from which spatio-temporal features were extracted by calculating the mean amplitudes in 10 discriminant time intervals, and classified using linear discrimination analysis (LDA) classifier to predict a target character of a 36 symbol ERP speller [19]. Moreover, CCA was used to detect four SSVEP stimuli in the form of four frequencies (5.45 Hz, 6.67 Hz, 8.57 Hz and 12 Hz) for SSVEP BCI system [20]. [21] investigated the feasibility of adding samples from different days to training set to improve the generalization of the emotions classifier, whereby five sessions in five different days were recorded for each subject. in this case training sample having the same number came from day 1,2,3 or day 4. The intervals between two consecutive sessions of a subject were randomly selected from four intervals (one day apart, two days apart, one week apart and two weeks apart) [21]. Consequently, an average classification accuracy of 64.9%, 68.7%, 70.9% and 73% for day 1, 2, 3 and day 4 was achieved respectively. To achieve the objectives of the investigation EEG signals were transformed to independent components, from which power spectral density was estimated for six frequency bands (theta, alpha, beta1, beta2, delta1 and delta2), then SVM was used to classify three emotion states (neutral, positive and negative) [21], [22], [23].

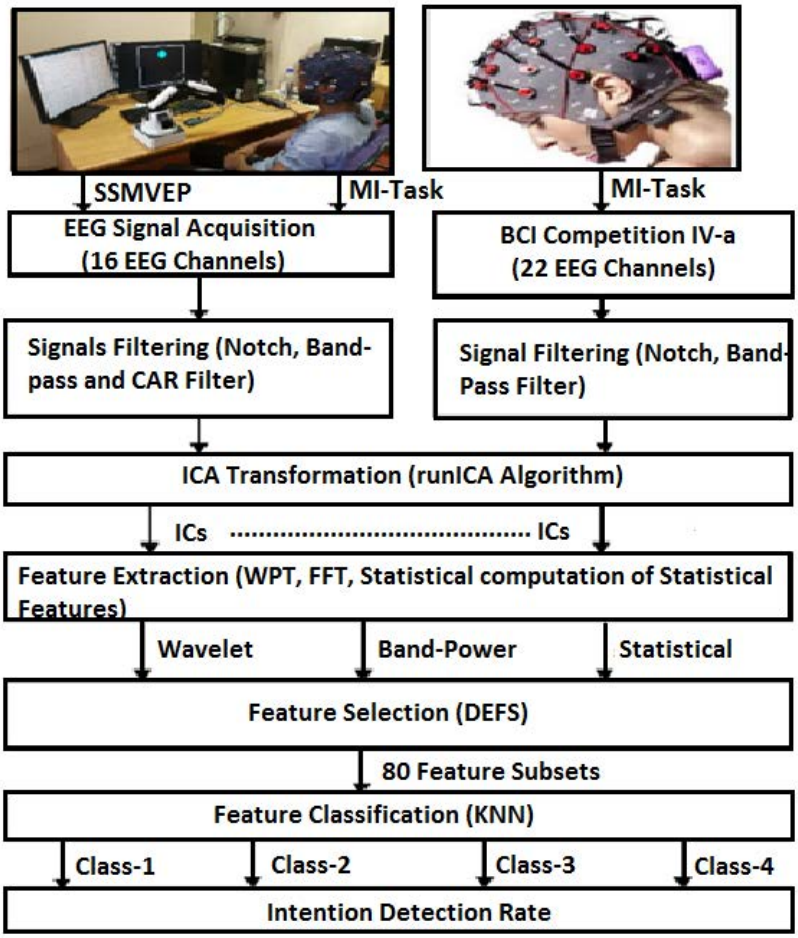


FIGURE 1. Proposed techniques for Multiclass EEG-based BCI framework.

In this study, we investigate five factors that constitute to low intention detection rate across sessions and subjects. The factors include concentration level, selection of ICs, inter-session variability, inter-subject variability and classification methods. Moreover, domain transfer learning approach is proposed to address the challenge of variability in EEG dynamics. As such transferability of neural information across multiple subjects/sessions is evaluated by comparing TL with two machine learning algorithms (NB and K-NN).

Firstly, three filters (notch, band-pass and common average reference (CAR)) were applied on three EEG datasets. Filtered EEG data was then transformed into independent components (ICA) using runICA algorithm. Statistical computation, Wavelet packet transform (WPT) and fast fourier transform (FFT) algorithm were used to extract features. In this case wavelet, band-power and statistical features were extracted from ICs. Relevant features were then selected from extracted feature sets using differential evolution feature selection (DEFS) algorithm. Furthermore, NB and K-NN classifier were used to predict four MI and SSMVEP classes respectively.

The rest of the paper is organized as follows. Section II gives a detailed description of methods used to achieve the objectives of this study namely EEG signal acquisition, signal pre-processing, feature extraction, feature selection and feature classification. Section III gives a detailed description of experimentation procedures. Section IV illustrates results of IC selection, concentration level, inter-session variability, inter-subject variability and classification methods experimentations. Section V discusses results for all five experimentations. Finally, some conclusions are provided in Section VI.

II. CLASSICAL BCI IMPLEMENTATION

In this paper, we address the challenge of variability in EEG characteristics that constitute to low intention detection rate in BCI. As popularly used, the raw EEG signals firstly go through independent component analysis (ICA). Five factors including changes in concentration level, selection of ICs, Feature classification, inter-session and inter-subject variability are investigated. The first experiment investigates the impact of concentration level on intention detection

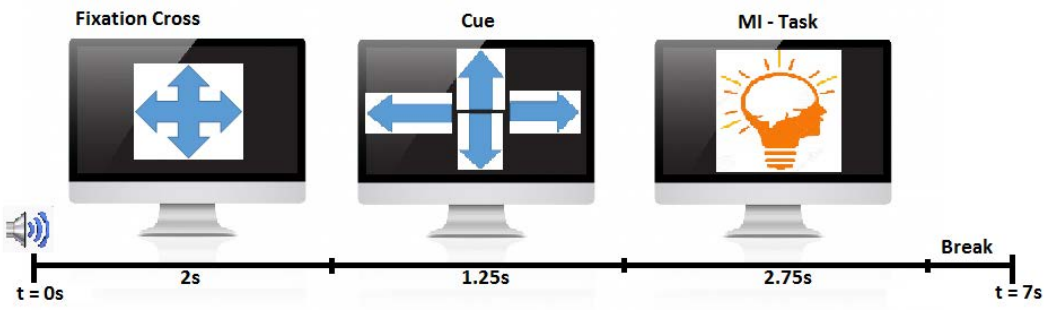


FIGURE 2. Experimentation paradigm timing scheme (BCI Competition IV-a).

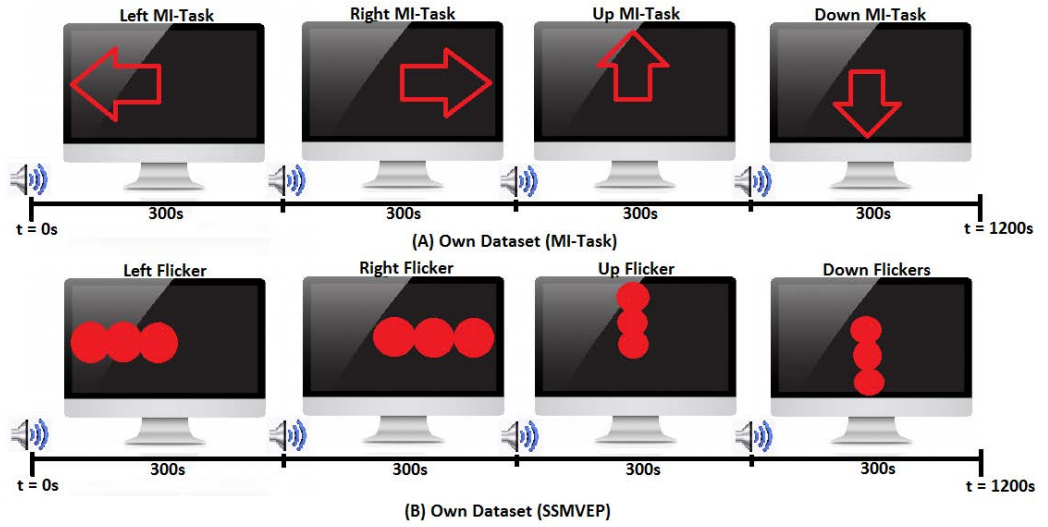


FIGURE 3. Experimentation paradigm timing scheme (Own dataset).

rate across sessions. To quantify the effect of concentration level, the relationship between the alpha rhythm and success rate is evaluated. In the second experiment we investigate whether IC selection can have an impact on intention detection rate across sessions. Then, the third experiment investigates whether there are similarities on features between ICs acquired from different sessions in different days. Consequently, samples from the first dataset ICs are used for training and validation, while the second dataset ICs are used for testing. Furthermore, the possibility of two subjects having similar EEG characteristics is investigated in the fourth experiment. In this case samples from one subject are used for training and validation, while samples from another subject are used for testing. Lastly, TL is compared with both K-NN and NB classifier mainly to determine whether neural information from multiple sources can be used to improve classification accuracy of individual sessions, subjects or domains.

A hybrid EEG based BCI system is proposed to facilitate the investigation as illustrated in Figure 1. As such three EEG datasets (BCI competition IV-a, our own MI-SSMVEP datasets) are used to investigated the above mentioned factors [2], [24]. In this case our own datasets were filtered using (50 Hz notch, 0.5-60Hz band-pass and CAR filter), while BCI competition IV-a dataset was pre-filtered using (50 Hz notch and 0.5-100 Hz band-pass) [25], [26]. Moreover, runICA algorithm in BCI2000 was then used to transform EEG signals into 16 and 22 ICs respectively [27]. Three sets of features (wavelet, band-power and statistical features) are extracted from artifact free components, whereby statistical computation, WPT and FFT algorithms were used [5]. DEFS algorithm was then used to select relevant features with high predictive capacity, while NB and K-NN classifier were used to predict both SSMVEP and MI classes respectively [23].

A. EEG SIGNAL ACQUISITION

1) BCI COMPETITION IV DATASET II-A

BCI competition IV is a publicly available database, consisting of EEG dataset II-a acquired in a controlled environment [28]. The dataset consists of four MI classes (left, right, both feet and tongue) recorded from the brain using twenty-two Ag/AgCl electrodes. EEG signals were recorded at a sampling rate of 250 Hz [29]. In this case electrodes were positioned on the scalp with a distance of 3.5cm apart using a 10-20 position system. A fixed cross '+' was displayed on

the screen at $t = 0$s to signify a start of a trial, while subjects were seated facing a projecting computer screen. An external stimulus in a form of a beeping sound after two seconds ($t = 2$s) was also used at the beginning of a trial. A visual cue represented by an arrow pointing to four directions was then displayed for $t = 1.25$s on the screen, from $t = 3.25$s until $t = 6$s subjects were required to perform motor imagery task as illustrated on the experimentation paradigm in Figure 2 [29].

2) OUR OWN RECORDED EEG DATASETS

Two of our own EEG datasets were recorded using a g.tec EEG recording system from five subjects, all participants had no BCI training prior to the experiment. In this case sixteen electrodes positioned on the scalp according to a 10-20 positioning system were used [19]. Both MI and SSMVEP EEG signals were sampled at 250 Hz. The first datasets consisted of four EEG classes (left hand, right hand, both feet and tongue) corresponding to four imagined movements [30]. At the beginning of a trial subjects were seated facing a projecting computer screen, and requested to perform four MI tasks for 300 seconds each. At the beginning of every 300 seconds an arrow corresponding to four MI tasks was displayed on the screen, the arrow was used to notify the subjects of which MI task to perform. A beeping sound was also utilized to notify the subjects to begin performing the MI task as illustrated in Figure 3(A).

Furthermore, the same experiment setup used during acquisition of MI-task was used for SSMVEP classes. In this case, at the beginning of the experiment, visual stimulus in the form of four flashing colored balls were displayed on the screen. The flashing balls were tagged with four frequencies (29 Hz, 13.3 Hz, 17 Hz and 21 Hz) representing four EEG classes [31]. Moreover, the flashing balls were moving in four directions corresponding to four movements (left, right, up and down). In a trial, subjects were requested to focus their attention on the flashing ball. In this instance a beeping sound was utilized to notify the subject to begin performing SSMVEP tasks for 300 seconds as illustrated in Figure 3(B). Acquisition of both MI-SSMVEP EEG datasets lasted for twenty minutes equivalent to a signal length of 300000 samples per epoch [32].

B. EEG SIGNAL FILTERING

1) EEG SIGNAL FILTERING

BCI competition IV-a dataset was pre-filtered using a 0.5 Hz to 100 Hz band-pass filter to eliminate the impact of non-physiological artifact such as noise, while a 50 Hz Notch filter was utilized to eliminate the impact of line noise [29].

$$|II(\omega)|^2 = \frac{1}{1+\omega^{2n}} \tag{1}$$

Our own EEG datasets were filtered using a 0.5 Hz to 60Hz band-pass filter. Equation (1) was used to compute a butterworth band-pass filter, with angular frequency denoted by ω in radian per second, and the filter order denoted by n [33]. Moreover, a notch filter at a cut-off frequency of 50 Hz was also applied on both datasets [29], [34].

$$V_i^{CAR} = V_i^{ER} - \frac{1}{n}\sum_{j=1}^{n} V_j^{ER} \tag{2}$$

Furthermore, common average reference (CAR) was applied to eradicate the impact of noise from electrodes, and improve signal-to-noise ration of EEG signals. In this case CAR was computed using (2), whereby the potential between ith EEG channel and reference is denoted by V_i^{ER}, while the number of electrodes in a montage is denoted by n [35].

2) EEG SIGNAL DECOMPOSITION

ICA algorithm was applied on all three EEG datasets to eliminate the impact of physiological artifacts such as electromyography (EMG), electrocardiogram (ECG) and electrooculogram (EOG) [25], [36]. Transforming EEG signals into ICs in order to select artifact free components is another way of addressing the challenge of interferences [36]. In this case artifactual components are rejected, while artifact free ICs are reconstructed into clean EEG signals [37]. Multivariate EEG signals in this instance were transformed into signals containing mutually independent components known as ICs [38].

$$x = As(t) = \sum_{i=1}^{m} a_i s_i \tag{3}$$

Equations (3) was used to compute ICs, whereby the number of observations containing random noises is represented by $x = (x_1, x_2, \ldots x_n)^T$, while $s = (s_1, s_2, \ldots s_n)^T$ represents the source signals. Moreover, the mixed matrix is denoted by A while a_i represents a vector of rows [25], [39].

$$z = Vx = WAs \tag{4}$$

Equation (4) represents the number of dimensional vector of independent components denoted by z achieved through a reversible matrix W, whereby V represents a whiten matrix [39].

C. FEATURE EXTRACTION

After signal decomposition, three sets of signal features (namely band-power, statistical and wavelet features) are extracted from ICs for each of the three datasets. As such, an extension of discrete wavelet transform (DWT) in the form of wavelet packet transform (WPT) [40], which is highly significant for exhibiting more informative high and low frequencies of a signal, was used to extract 255 wavelet features [40].

$$x(t) = \sum_{k=-\infty}^{+\infty} C_{N,K}\theta\left(2^{-N}t - k\right) + \sum_{j=1}^{N}\sum_{k=-\infty}^{+\infty} d_{j,k}2^{\frac{-j}{2}}\varphi(2^{-j}t - k) \tag{5}$$

In this case, WPT was computed using (5), in which a daubuchies of order 4 (db4) mother wavelet was utilized to transform ICs into seven decomposed levels or wavelet packet tree [41], [42]. The output from each decomposed level represents approximation and detail coefficients. A wavelet function denoted by $\varphi(t)$ was used to decompose EEG signals into detail coefficients, while a scaling function denoted by $\theta(t)$ was used to decompose signals into approximation coefficients. In this case N representing the number of samples acquired from the source signals. J represents the number of scales, while the subband index within the scale is represented by k. The coarse and detail level expansion coefficients is denoted by $C_{N,K}$ And $d_{j,k}$ respectively. The dilated and translated versions of the scaling and wavelet function is denoted by $(2^{-N}t - k)$ and $(2^{-j}t - k)$ respectively [9], [43]. Subsequently, a sliding window approach was applied on a wavelet packet tree, whereby wavelet coefficients were separated into multiple windows from which wavelet features are extracted. Each window size was set to 750, while each window was incremented by 750. As such a single wavelet feature was extracted in the first decomposition level, while the number of extracted features in each level doubled until the seventh level, mainly because of wavelet coefficients in each decomposition level. In this case the decomposition levels of a wavelet packet tree are responsible for the resulting 255 feature size.

$$Y(k) = \frac{1}{N}\int_{i=0}^{N-1} y(t)e^{-j2\pi\left(\frac{k}{N}\right)i} \tag{6}$$

Furthermore, seven EEG frequency bands (delta, theta, alpha, mu, central beta, beta and gemma) representing band-power features, were extracted from ICs using FFT [35], [44], [45]. Ten statistical features (namely mean, median, standard deviation, mean absolute deviation, skewness, kurtosis, spectral entropy and dominant frequency characteristics (maximum frequency, maximum value, maximum ratio)) were also extracted using FFT [46]. As such, (6) was used to compute fast Fourier transform (FFT), with $y(t)$ representing sampled values, while the size of a signal domain is denoted by N and i represents indices of a domain vector [35], [44], [45].

D. FEATURE SELECTION

Differential-evolution-based channel and feature-selection (DEFS) algorithm was used to eradicate the problem of dimensionality and redundant features. This technique detects more informative feature subsets, consisting of high predictive power [47]. DEFS algorithm uses differential-evolution (DE) optimization and repair mechanism to select the best feature subsets [47]. Subsequently (7) defines a scale factor F utilized to control iterations between selected feature population, c_1 representing a constant lower than 1 [47].

$$F = \frac{\mathrm{c_1 rand}}{\max(x_{j,r1,\mathrm{g}}, x_{j,r2,\mathrm{g}})} \tag{7}$$

Equation (8) was utilized to enable population members to oscillate within bounds without crossing the optimal solutions, with NF representing the number of features [47].

$$x_{j,i,\mathrm{g}} = \begin{Bmatrix} \mathrm{NF} & \text{if } x_{j,i,\mathrm{g}} > \mathrm{NF} \\ 1 & \text{if } x_{j,i,\mathrm{g}} < 1 \end{Bmatrix}. \tag{8}$$

Equation (9) representing a distribution factor $FD_{j,g}$ was utilized to prevent selected features from being selected more than once in the same feature vector, in which the total number of features is denoted by NF. Suitably chosen positive constant is denoted by a_1 that shows the significance of features in PD. In this case PD_j represents subsets with a lower fitness as compared to the average fitness of the entire population, while DNF represents the desired number of features to be selected, and ND_j representing subsets with a higher fitness as compared to the average fitness of the entire population [47].

$$\mathrm{FD}_{j,\mathrm{g}} = a_1\left(\frac{\mathrm{PD}_j}{\mathrm{PD}_j + \mathrm{ND}_j}\right) + \frac{\mathrm{NF} - \mathrm{DNF}}{\mathrm{NF}}\left(1 - \frac{(\mathrm{PD}_j + \mathrm{ND}_j)}{\max(\mathrm{PD}_j + \mathrm{ND}_j)}\right) \tag{9}$$

Equation (10) was utilized to compare the previous iteration to the current iteration to determine features that have made substantial improvement, and grant higher weights to improved features which are then utilized in the next iteration. In this case FD represents distribution factor [47].

$$T = \left(\mathrm{FD}_{\mathrm{g}+1} - \mathrm{FD}_{\mathrm{g}}\right)\mathrm{FD}_{\mathrm{g}+1} + \mathrm{FD}_{\mathrm{g}} \tag{10}$$

Equation (11) was utilized to determine the number of times a specific feature was utilized within each iteration based on the updated distribution factor [47].

$$T = T - 0.5\mathrm{rand}\,(1, \mathrm{NF})\,(1 - T) \tag{11}$$

Consequently, several relevant parameters were assigned for the DEFS algorithm to select relevant feature subsets. The desired number of features (DNF) was set to 80, and the population size (PSIZE) was set to 150, while the number of generations (GEN) was set to 1000 as the terminating condition [47].

E. FEATURE CLASSIFICATION

A transfer learning approach is proposed in this section and compared with two supervised machine learning algorithms namely K-NN and NB, mainly to investigate whether multi-source neural information can improve the classification accuracy of individual sessions or subjects.

1) K-NEAREST NEIGHBOR (K-NN)

A supervised machine-learning algorithm was used to classify both MI and SSMVEP classes acquired from each of the three datasets respectively. K-nearest neighbor (K-NN) was used to classify samples by locating k samples containing similar independent variables in the training set [4], [43].

$$D_E(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{12}$$

To obtain different set of k nearest neighbors, K-NN calculate the distance across instances, in which the Euclidean distance is used to obtain the distance between instances. Equation (12) was used to compute Euclidean distance, whereby two vectors in the feature space are denoted by x and y, and their coordinates is denoted by x_i and y_i respectively [47], [48].

2) NAÏVE BAYES (NB)

A statistical technique based on the Bayes theorem know as naïve bayes (NB) was used to predict the probability of a class-generating-observed value for features, it assumes that an individual feature of a certain class is isolated from other features [7].

$$c\,(E) = \text{argmax}_{c\in C}\; P(c)\prod_{i-1}^{n} P(a_i|c) \tag{13}$$

Naïve Bayes (NB) is defined by equation (13), whereby E signify a vector $<a_1, a_2, \ldots, a_n>$, a_i representing attributes, class variables denoted by C, c representing the value of C and the class labels from which E belongs to represented by $c\,(E)$ [7].

3) TRANSFER LEARNING

In this section, an inductive transfer learning approach is proposed to classify EEG signals acquired from various sources. TL in this case compensate for minimal labeled data by using data represented in different feature space to match the distribution between both source and target domains [49]. Subsequently the proposed approach uses labeled data from multiple source domains and labeled data from a single target domain to train a model [50]. As such neural information from different multiple sources or source domains are utilized to enhance the learning performance of the target domain [51]. In this case data from a single session\subject is considered as a target domain, while data from multiple EEG sources are considered as source domains.

III. EXPERIMENTAL SETUP

A. INTER-SESSION VARIABILITY

In this study, five factors are investigated in session-to-session experiments. All own datasets were acquired in different days, consisting of a single session each. In this case, each MI or SSMVEP task was performed for 5 minutes each, equivalent to 100 three-second windows or 75000 samples. All sessions were 20 minutes long which is equivalent to 400 three-second windows or 300000 samples [52], [53]. The IDR of each session was obtained by summing the IDRs across 400 three-second windows. In this case each accurately predicted 3 second window was considered as a 1, while each misclassified 3 second was considered as a 0, then the results for 400 three-second windows were summed up to give an overall classification accuracy. Moreover, the alpha rhythm across each session was obtained by averaging 100 alpha rhythm peaks [21].

B. CONCENTRATION LEVEL

In this experiment, the alpha rhythm denoted by a frequency band (8Hz – 13Hz) is used as an indicator of concentration level. In this case FFT algorithm is used to extract the alpha rhythm within a frequency between 8Hz and 13Hz from each three second window of EEG datasets. As such intention detection rate (IDR) and the alpha rhythm for each three second window is compared, mainly to evaluate the relationship between the alpha rhythm and IDR. In a case where the alpha rhythm is high and the corresponding IDR is low then the concentration level is low.

C. ICA COMPONENTS MERGE

IC selection experiment was extended to investigate whether ICs acquired in different sessions from the same subject possess similar EEG characteristics. As such samples from IC in the first session (Se1(IC-9)) were used for training and validation, while samples from the same IC in the second session (Se2(IC-9)) were used for testing during feature classification [14], [15]. In this case, the same procedure was applied for the remaining two conditions (IC-10(Se1-Se2) and IC-11(Se1-Se2)).

To be noted that ICA does not have a ordering/sorting mechanism, therefore the order of an ICA component does not have a physical meaning. In this paper, we used the specific ICA components to discuss their impacts on the EEG-BCI performance, and mainly intended to demonstrate that different ICA component selection could lead to different performance, therefore, show that ICA component selection is one factor to be considered.

D. INTER-SUBJECT VARIABILITY

The fourth experiment investigates whether two different subjects possess similar EEG characteristics. Consequently, samples from one IC acquired from one subject (S1(IC-9)) were used for training and validation, while samples from the same IC acquired from another subject (S2(IC-9)) were used for testing during feature classification [14], [15]. The same procedure was repeated for the remaining two conditions (IC-10(S1-S2) and IC-11(S1-S2)).

E. FEATURE CLASSIFICATION

In this experiment transfer learning is compared with K-NN and NB, mainly to investigate whether knowledge transfer from multi-source domains can enhance performance of classification algorithms. As such transfer learning approach uses knowledge from source domains to improve learning performance of the target domain [54]. In this case when transfer learning approach is applied on each dataset, one domain (subject\session) is considered as a target domain, while the remaining domains (subjects\sessions) are treated as source domains. Moreover, when K-NN and NB classifiers are applied on each dataset respectively, samples from multiple sources (subjects\sessions) are utilized to train

the classifiers, while samples from a single subject\session are used to test the classifiers respectively.

IV. EXPERIMETAL RESULTS

A. FACTORS TO BE ANALYSED

1) FACTOR 1: SELECTION OF ICA COMPONENT

In this section we investigate whether IC selection can significantly affect the performance of both MI and SSMVEP based BCI. Firstly, 16 EEG channels were transformed by ICA to get 16 ICs, which is a linear mixture of multiple EEG channels [38]. As such, the IDRs across sessions of all 16 ICs were evaluated and compared. In this case Se1(IC11) achieved the highest IDR of 100% in SSMVEP Se1, while in Se2, the same IC (IC11) achieved an IDR of 83% only, as illustrated in TABLE 1. Interpreted to a 17% drop in accuracy was observed across Se1 and Se2. Moreover, Se2(IC7) achieved the highest of 92% in the second SSMVEP session, while Se1(IC7) achieved an accuracy of 78%, a 14% drop in accuracy was observed across both sessions. A 35% increase in accuracy across Se1(IC15) and Se2(IC15) was observed, which is the highest variation across both sessions, 53% and 88% in the two SSMVEP sessions respectively. Second to this, Se1(IC2) achieved the lowest accuracy of 51%, but 80% in Se2, meaning a 29% variation in IDR. Both positive and negative variations were observed in TABLE 1.

For MI sessions, Se2(IC2) and Se2(IC13) achieved the highest accuracy of 69%, while Se1(IC2) and Se1(IC13) achieved an accuracy of 57% and 48% only, as illustrated in TABLE 1. In this instance a 12% variation in IDR occurred across Se1(IC2) and Se2(IC2), while a 21% difference was observed across Se1(IC13) and Se2(IC13). Moreover, Se1(IC9) achieved the lowest accuracy of 40%, while Se2(IC9) achieved an accuracy of 62% in MI session 2, meaning a 22% variation.

From TABLE 1, one finds that individual ICs for SSMVEP sessions yielded significant variation in IDR, from component-to-component and from session-to-session. In this case, a significant decline in accuracy between sessions was observed across components (IC2, IC5, IC8, IC10, and IC15). Moreover, ICs in both MI sessions in most cases achieved a high success rate, however it varies from IC-to-IC. Some significant variations across sessions were observed (in this case across IC4, IC9, and IC13). Based on these results we can conclude that selection of ICs does significantly affect the performance of both MI and SSMVEP based BCI.

2) FACTOR 2: CONCENTRATION LEVEL

The alpha rhythm denoted by a frequency band (8Hz - 12Hz) is responsible for basic functions (such as concentration) or neural processes of the brain [55]. Subsequently, the alpha

TABLE 1. IDRs for individual components.

ICA Components	SSMVEP Session 1 (Se1)	SSMVEP Session 2 (Se2)	MI Session1 (Se1)	MI Session 2 (Se2)	Average (SSMVEP Sessions)	Average (MI Sessions)
IC1	72%	85%	48%	59%	79%	54%
IC2	51%	80%	57%	69%	66%	63%
IC3	63%	59%	49%	62%	61%	56%
IC4	82%	78%	43%	64%	80%	54%
IC5	94%	74%	57%	63%	84%	60%
IC6	73%	81%	57%	64%	77%	61%
IC7	78%	92%	54%	63%	85%	59%
IC8	84%	64%	42%	47%	74%	45%
IC9	79%	91%	40%	62%	85%	51%
IC10	86%	62%	47%	61%	74%	54%
IC11	100%	83%	55%	58%	92%	57%
IC12	96%	88%	52%	55%	92%	54%
IC13	79%	82%	48%	69%	81%	59%
IC14	91%	86%	51%	49%	89%	50%
IC15	53%	88%	62%	47%	71%	55%
IC16	61%	80%	65%	64%	71%	65%

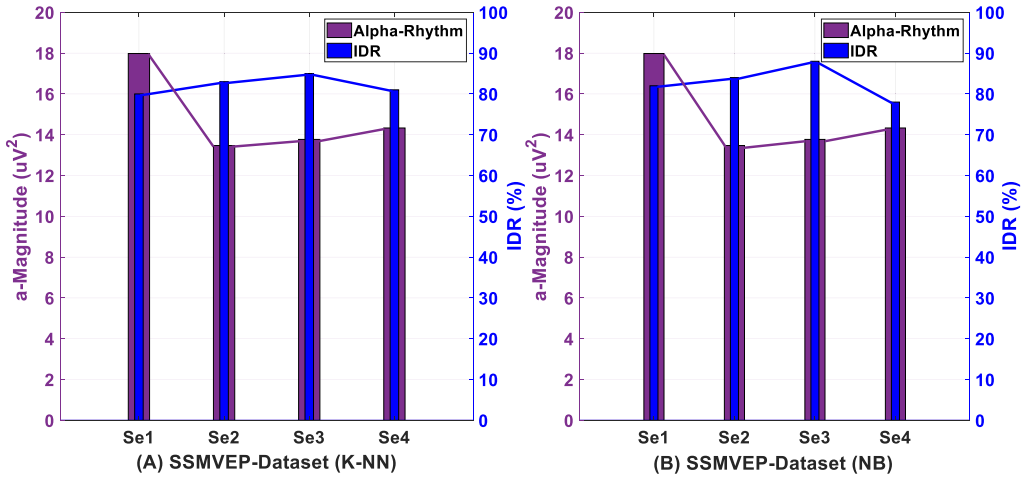


FIGURE 4. Relationship between alpha rhythm and success rate (SSMVEP).

rhythm magnitude is used as an indicator of change in concentration level [45]. SSVEP signals are mainly dependent on frequency tags flickering at different frequencies [56].

In this section, we investigate whether change in concentration level across sessions can affect the performance of SSMVEP based BCI. To facilitate the investigation our own SSMVEP dataset is used in this experiment. The alpha rhythm and IDRs obtained from 100 three-second windows or 75000 samples in a single session were summed and averaged respectively. Concentration level is hard to measure. Fortunately, it can be represented by the strength of alpha rhythm. It was commonly accepted that the higher alpha rhythm indicates a low concentration level, while a weaker alpha rhythm means a higher concentration level. So in this experiment, we study the relationship between the alpha rhythm magnitude and the IDR, instead of using the concentration level directly.

As shown in Figure 4, the negative correlation between the IDR using K-NN and alpha rhythm magnitude is clear. The highest alpha rhythm at a magnitude of 17.98 uV^2 was obtained in the first session (Se1) and associated to 80% IDR, when the alpha rhythm magnitude decreased to 13.47 uV^2 in the second session (Se2) the IDR increased to 83%, as illustrated in Figure 4(A). In the third session (Se3) an alpha rhythm magnitude of 13.77 uV^2 is linked to 85% IDR. A 14.33 uV^2 alpha rhythm is observed in the fourth session (Se4) and associated to 81% IDR.

The same process was repeated using NB classifier as illustrated in Figure 4(B). In this case an alpha rhythm magnitude of 17.98 uV^2 associated to 82% IDR in the first session (Se1), while an alpha rhythm magnitude decreased to 13.47 uV^2 and IDR increased to 84% in the second session (Se2) comparing to Se1. In the third session (Se3) an alpha rhythm magnitude of 13.77 uV^2 is linked to 88% IDR. A 14.33 uV^2 alpha rhythm magnitude is observed with a 78% IDR in the fourth session (Se4).

From these experiments, one finds that higher alpha rhythms (representing a lower concentration levels) were generated in Se1 and Se4, and resulted in lower IDRs. However, lower alpha rhythm magnitudes corresponding to higher concentration level were observed in Se2 and Se3, which resulted in higher IDRs. This means the existence of negative correlation between the concentration level and IDR for both SSMVEP and MI based BCIs. Therefore, the concentration level is a significant factor to the IDR.

These experiments also reveal that the SSMVEP is not mainly relevant to the frequency of flashing as commonly recognized, but also affected by concentration levels of the users.

Motor imagery (MI) task is a mental process in which a subject performs imagined limb movements, as such MI task modulation is highly dependent on concentration [39]. This section investigates if the concentration levels across sessions affect the IDR of MI based BCIs. Our own recorded MI dataset was used to facilitate the investigation. As such the highest alpha rhythm magnitude of 232.39 uV^2 associated to 49% IDR was observed in the third session (Se3) using K-NN classifier, while the lowest alpha rhythm magnitude of 202.824 uV^2 is linked to 62% IDR in the second session (Se2), as illustrated in Figure 5(A). Moreover, slightly similar alpha rhythm magnitudes of 222.04 uV^2 and 222.47 uV^2 were observed in the first (Se1) and the fourth (Se4) sessions, where similar IDRs were achieved. Both the lowest alpha rhythm magnitude and the highest IDR were captured in Se2.

Moreover, NB classifier was further employed to evaluate the relationship between the alpha rhythm magnitudes and IDRs in MI sessions as illustrated in Figure 5(B). Consequently, the alpha rhythm magnitude of 232.39 uV^2 associated to 48% IDR was observed in the third session (Se3), while the lowest alpha rhythm magnitude of 202.824 uV^2 is linked to 60% IDR in the second session (Se2). Moreover,

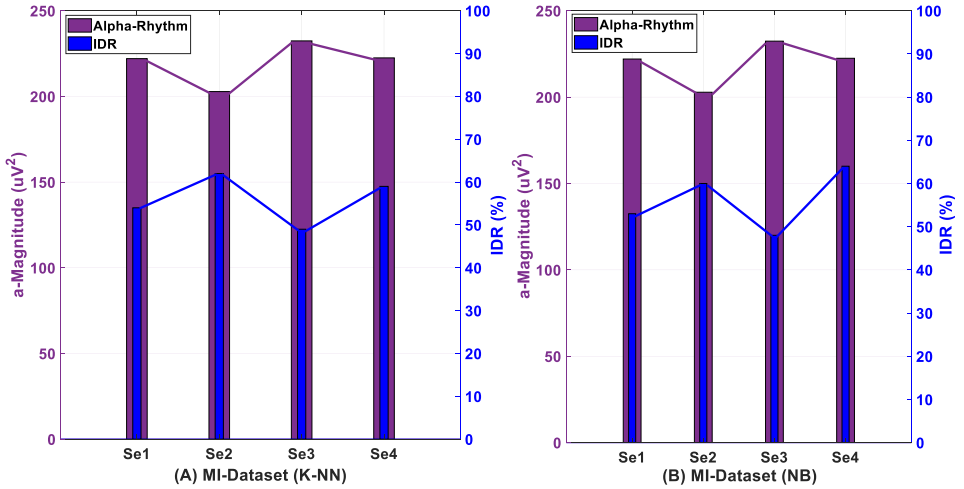


FIGURE 5. Relationship between alpha rhythm and success rate (MI).

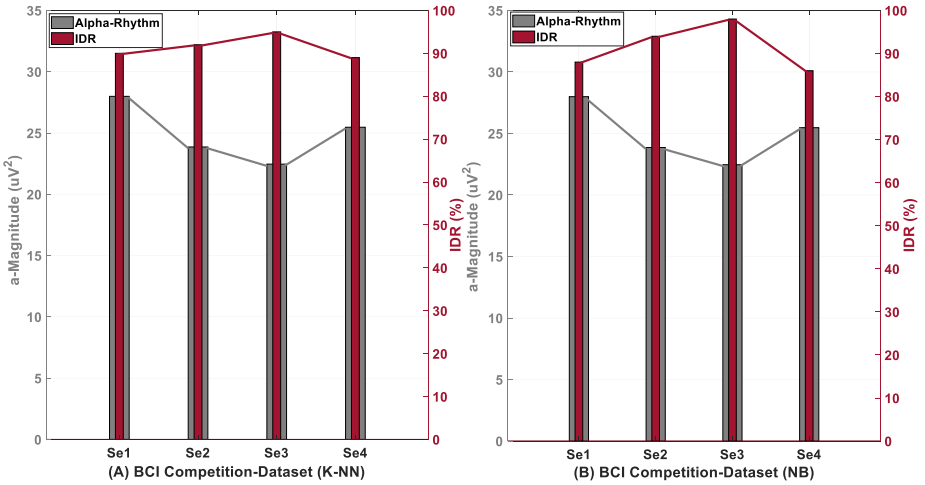


FIGURE 6. Relationship between alpha rhythm and success rate (BCI Competition IV-a (MI)).

slightly similar alpha rhythm magnitudes of 222.04 uV^2 and 222.47 uV^2 were observed in the first (Se1) and the fourth (Se4) sessions, and resulted in IDRs of 53% and 64% respectively.

These experiments show the negative correlation between the alpha rhythm magnitudes and IDRs, therefore, one can conclude that the concentration levels affect the IDR of MI tasks.

The effect of changes in concentration levels on the IDRs of MI based BCI is further investigated using BCI competition IV-a datasets. The relationship between the alpha rhythm magnitudes and the IDRs across sessions is evaluated, as shown in Figure 6. The highest alpha rhythm magnitude of 27.99 uV^2 is associated to 90% IDR in the first session (Se1), while after alpha rhythm magnitude decreased to 23.86 uV^2 in Se2 the IDR increased to 92% when K-NN is employed as illustrated in Figure 6(A). Furthermore, when the alpha rhythm further decreased to 22.46 uV^2 in Se3, then the IDR increased to 95%. When NB classifier is employed an alpha rhythm magnitude of 27.99 uV^2 is associated to 88% IDR in the first session (Se1), while after alpha rhythm magnitude decreased to 23.86 uV^2 in Se2 the IDR increased to 94% in the as illustrated in Figure 6(B). Furthermore, if the alpha rhythm further decreased to 22.46 uV^2 in Se3, then the IDR increased to 98%.

From these experiments, one finds that a lower alpha rhythm magnitude corresponds to a higher concentration level denoted by a higher IDR. This further validates the negative correlation between concentration levels and IDRs for the MI tasks.

3) FACTOR 3: INTER-SESSION VARIABILITY

Diverse mental states as a result of both known and unknown factors give rise to significant challenges in BCI, namely session-to-session variability suspiciously constituting to variations of IDR. Following this hypothesis, the

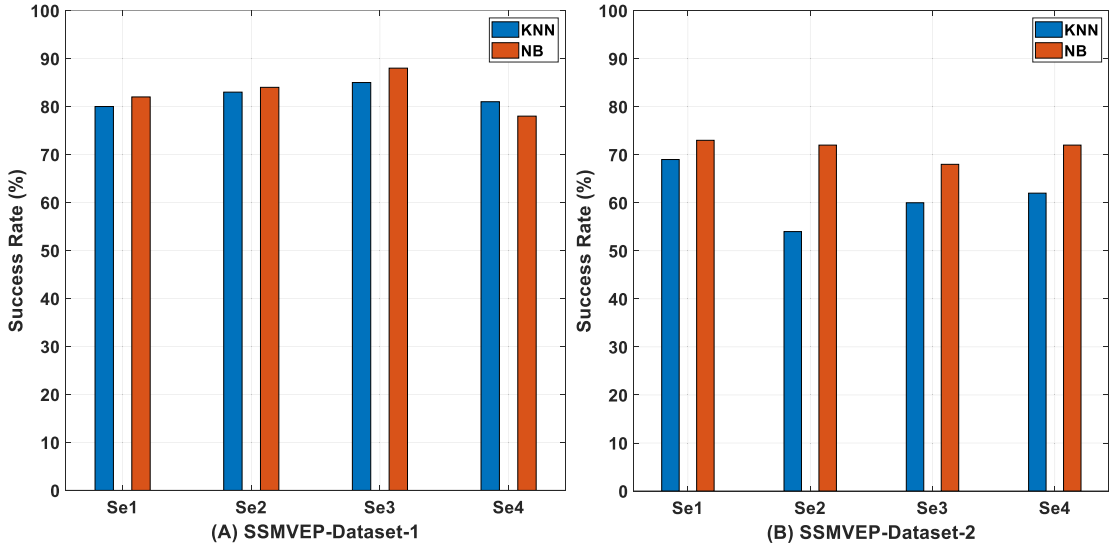


FIGURE 7. Session-to-session variability results (SSMVEP).

impact of session-session variability on SSMVEP based BCI performance is investigated in this section.

Two datasets (SSMVEP dataset-1 and SSMVEP dataset-2) recorded for the same subject but in different days are used to facilitate the investigation, in which the IDRs across inter-sessions and intra-sessions are evaluated. SSMVEP dataset-1 is used to extract the features, train and test two classifiers (K-NN and NB). Similarly, SSMVEP dataset-2 is used to extract the features, train and test two classifiers (K-NN and NB). In this case an average IDR of 82% and 83% is achieved in the first dataset, while an average IDR of 61% and 71% in the second dataset for both K-NN and NB classifier respectively. However, the IDR variation across sessions due to multiple factors resulted in a 21% drop in IDR across both datasets when K-NN classifier is employed, while a 12% drop in IDR across both datasets was observed when NB classifier is employed. The variation of IDRs among sessions of both datasets can significantly affect the average success rate. In this case Se1 and Se4 achieved a lowest accuracy of 80% and 78%, while Se3 achieved a highest accuracy of 85% and 88% across all four sessions in SSMVEP dataset-1 for both K-NN and NB classifiers respectively, i.e., a highest decline in accuracy of 5% was observed for K-NN between Se1 and Se3, while 8% decline in accuracy was observed between Se3 and Se4 for NB classifier as illustrated in Figure 7(A). Furthermore, Se1 achieved a highest accuracy of 69% and 73%, while Se2 and Se3 achieved a lowest accuracy of 54% and 68% across all four sessions in SSMVEP dataset-2 for both K-NN and NB classifiers respectively. Consequently, a highest decline in accuracy of 15% was observed between Se1 and Se2 for K-NN classifiers, while 4% decline across Se2 and Se3 was observed for NB classifiers as illustrated in Figure 7(B). The first dataset in this instance yielded significant results as compared to the second dataset. Moreover, a higher decline in accuracy was observed when K-NN classifier is employed as compared to NB classifier. However, IDRs vary across sessions and same subject datasets acquired in different days. Based on these results we can conclude that session-to-session variability does affect the performance of SSMVEP based BCI [10], [14], [16].

This section further investigates the effect of session-to-session variability on BCI performance with a focus on MI task. Two MI datasets (MI dataset-1 and MI dataset-2) recorded for the same subject but in different days are used to facilitate the investigation, in which the IDRs across inter-sessions and intra-sessions are evaluated. MI dataset-1 is used to extract the features, train and test the classifier. Similarly, MI dataset-2 is used to extract the features, train and test the classifier. Subsequently, an average IDR of 56% is achieved in the first dataset, while an average IDR of 44%, and 52% in the second dataset for both K-NN and NB classifier respectively. However, the IDR variation across sessions due to multiple factors resulted in a 12%, and 4% drop in IDR across both datasets using K-NN and NB classifier respectively. In this case Se3 achieved a lowest accuracy of 49% and 48%, while Se2 and Se4 achieved a highest accuracy of 62% and 64% across all four sessions in MI dataset-1 using K-NN and NB classifier respectively, i.e., a highest decline in accuracy of 13%, and 12% was observed between Se2 and Se3 when both K-NN and NB classifier respectively as illustrated in Figure 8(A). Furthermore, Se2 achieved a highest accuracy of 47% and 54%, while Se4 achieved a lowest accuracy of 37% and 48% across all four sessions in MI dataset-2 for both KNN and NB classifier respectively. Consequently, a highest decline in accuracy of 9% and 5% was observed between Se3 and Se4 for both K-NN and NB classifier respectively as illustrated in Figure 8(B). Moreover, sessions in the first dataset yielded higher IDRs as compared to sessions in the

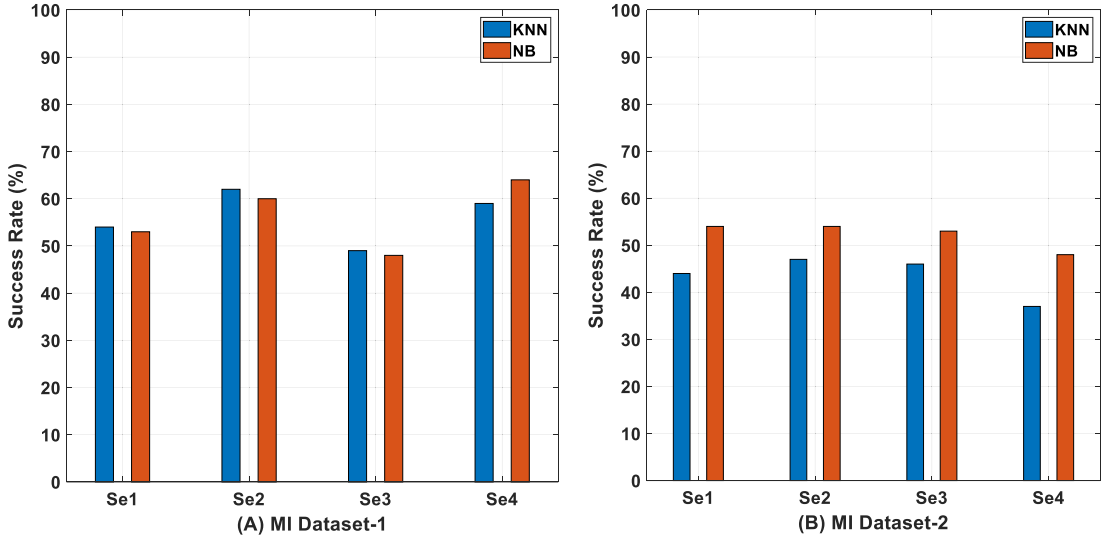


FIGURE 8. Session-to-session variability results (MI).

TABLE 2. Results for inter-session variability.

	K-NN						NB					
	SSMVEP		MI		BCI Competition		SSMVEP		MI		BCI Competition	
ICs	WSe (Se1-Se1)	CSe (Se1-Se2)	WSe (Se1-Se1)	CSe (Se1-Se2)	WSe (Se1-Se1)	CSe (Se1-Se2)	WSe (Se1-Se1)	CSe (Se1-Se2)	WSe (Se1-Se1)	CSe (Se1-Se2)	WSe (Se1-Se1)	CSe (Se1-Se2)
IC9	94%	65%	67%	59%	86%	35%	78%	39%	66%	50%	43%	40%
IC11	100%	45%	60%	46%	72%	40%	69%	55%	58%	46%	43%	32%
IC12	96%	60%	66%	57%	74%	40%	70%	51%	63%	58%	50%	43%
Average	97%	57%	64%	54%	77%	38%	72%	48%	62%	51%	45%	38%

second dataset. However, IDRs across sessions fluctuate in turn affect the overall success rate. As such we can conclude that session-to-session variability does affect the performance in the form of success rate of MI based BCI [16], [21].

EEG signals are non-linear and non-stationary, as such the magnitude of EEG signals change over time due to external and internal interferences. In this section, ICs selection experiment is extended, mainly to explore the feasibility of the same ICs acquired in two different sessions from the same subject to contain similar EEG characteristics. The most significant ICs (IC9 $\sim$ IC11) are used to investigate the inter-session variability. Samples from IC9 in the first session were used for training and validation, while samples from IC9 in the second session were used for testing. This procedure is denoted by CSe representing cross-sessions, while WSe represents within-session experiment. The same procedure (Se1-Se2) was repeated for IC10 and IC11 for all datasets. Furthermore, samples from IC-9 in a single session were used for training, validation and testing. The same procedure (Se1-Se1) was repeated for IC10 and IC11, for both SSMVEP and MI sessions. Consequently, (Se1-Se2) was compared with (Se1-Se1) to further validate the impact of inter-session variability.

In this case K-NN algorithm was used to classify both (Se1-Se1) and (Se1-Se2) samples as illustrated in TABLE 2. As such IC10(Se1–Se1) achieved a highest accuracy of 100%, while IC10(Se1–Se2) achieved lowest accuracies of 45% for SSMVEP dataset. A highest decline in accuracy of 55% was observed when samples from IC10(Se1) are integrated with samples from IC10(Se2). Moreover, a highest decline in accuracy of 14% for MI dataset was observed, when samples from IC11(Se1) are integrated with samples from IC11(Se2). In this case IC11(Se1–Se2) achieved a lowest accuracy of 46%, while IC11(Se1–Se1) an accuracy of 60%. A highest decline in accuracy of 51% was observed for BCI competition dataset, when samples from IC9(Se1) were merged with samples from IC9(Se2). In this instance IC9(Se1–Se2) achieved a lowest accuracy of 35%, while IC9(Se1–Se1) an accuracy of 86%.

Moreover, the same experiment was repeated using NB algorithm to classify both (Se1-Se1) and (Se1-Se2) samples as illustrated in TABLE 2. As such IC9(Se1–Se1) achieved a highest accuracy of 78%, while IC9(Se1–Se2) achieved lowest accuracies of 39% for SSMVEP dataset. A highest decline in accuracy of 39% was observed when samples from IC9(Se1) are integrated with samples from IC9(Se2).

TABLE 3. Results for inter-subject variability.

	K-NN						NB					
	SSMVEP		MI		BCI Competition		SSMVEP		MI		BCI Competition	
ICs	WS (S1-S1)	CS (S1-S2)	WS (S1-S1)	CS (S1-S2)	WS (S1-S1)	CS (S1-S2)	WS (S1-S1)	CS (S1-S2)	WS (S1-S1)	CS (S1-S2)	WS (S1-S1)	CS (S1-S2)
IC9	86%	30%	70%	35%	90%	44%	89%	45%	64%	46%	39%	36%
IC10	97%	38%	69%	48%	78%	28%	99%	48%	56%	41%	35%	31%
IC11	94%	65%	65%	41%	86%	34%	87%	74%	58%	45%	45%	34%
Average	92%	44%	68%	41%	85%	35%	92%	56%	60%	44%	40%	34%

Moreover, a highest decline in accuracy of 16% for MI dataset was observed, when samples from IC9(Se1) are integrated with samples from IC9(Se2). In this case IC9(Se1–Se2) achieved an accuracy of 50%, while IC9(Se1–Se1) a highest accuracy of 66%. A highest decline in accuracy of 11% was observed for BCI competition dataset, when samples from IC10(Se1) were merged with samples from IC10(Se2). In this instance IC10(Se1–Se2) achieved a lowest accuracy of 32%, while IC10(Se1–Se1) an accuracy of 43%.

Based on these results, it is worth noting that non-linearity and non-stationarity of EEG signals affects neural dynamics of subjects, resulting in significant variability across sessions of a single subject. In this case using samples from a single session for training, validation and testing the classifier yielded significant results. However, there was a significant drop in IDRs when samples from different sessions were merged. Moreover, samples from different component pair vary from session-to-session, as such BCI performance in the form of IDRs will in turn be affected.

4) FACTOR 4: INTER-SUBJECT VARIABILITY

Non-stationary characteristics of EEG signals constitute to existence of variability between subjects, and in turn affects IDR, when common features and classifiers are used across users, i.e. inter-subject scenario. This section investigates whether two different subjects can possess some similar EEG characteristics. Two of our own recorded MI and SSMVEP datasets are used to facilitate the investigation.

Samples from one subject are used for feature extraction, classifier training and validation, while samples from the other subject are used for testing. In this case, samples from IC9 acquired from the first subject, denoted as IC9(S1), are used for feature extraction, and classifier training and validation, because this IC performs well for this subject. Samples from IC9 acquired from the second subject, denoted as IC9(S2), are used for testing the classifier, while using the same method and parameters to extract features from IC9(S2) as IC9(S1) used. This procedure is denoted by CS representing cross-subjects, while WS represents within-subject experiment. The resulting IDRs are denoted as IC9(S1-S2) to present using IC9 of the first subject to classify the IC9 of the second subject. IC10 and IC11 are also considered due to the same reason. Therefore, the same procedure is repeated on IC10 and IC11, for both SSMVEP and MI dataset. In this case K-NN algorithm was used to classify both (S1-S1) and (S1-S2) samples.

TABLE 3 shows the results, from where one finds that IC11(S1-S2) achieved the highest IDR of 65% in the inter-subject experiments for SSMVEP dataset, while IC9(S1-S2) and IC10(S1-S2) achieved only 30% and 38% respectively. Moreover, IC10(S1-S2) achieved the highest accuracy of 48%, while IC9(S1-S2) and IC11(S1-S2) achieved 35% and 41% respectively for MI dataset.

Similarities in EEG characteristics in different subjects is further explored using BCI competition IV-a dataset for inter-subject experiments, where IC9, IC10, and IC11 are used as example to study the effects of inter-subject EEG characteristics on IDRs. TABLE 3 also shows the results, where IC9(S1-S2) achieved the highest accuracy of 44%, while IC10(S1-S2) and IC11(S1-S2) achieved 28% and 34% respectively.

Furthermore, the same experiment was repeated using NB algorithm to classify both (S1-S1) and (S1-S2) samples. As such TABLE 3 shows the results, from where one finds that IC11(S1-S2) achieved the highest IDR of 74% in the inter-subject experiments for SSMVEP dataset, while IC9(S1-S2) and IC10(S1-S2) achieved only 45% and 48% respectively.

Moreover, IC9(S1-S2) achieved the highest accuracy of 46%, while IC10(S1-S2) and IC11(S1-S2) achieved 41% and 45% respectively for MI dataset. Moreover, TABLE 3 also shows inter-subject results using BCI competition IV-a dataset, where IC9(S1-S2) achieved the highest accuracy of 36%, while IC10(S1-S2) and IC11(S1-S2) achieved 31% and 34% respectively.

Based on these results, it is worth noting that SSMVEP based BCI yielded higher inter-subject IDR, when IC11 (S1-S2) is used. However, there was a 29% decline in IDR when compared to IC11(S1-S1), in this case (S1-S1) represents training, validation and testing samples from a single subject. A 21% decline in IDR was observed when IC11(S1-S2) is compared with IC11(S1-S1), whereby IC11(S1-S1) achieved an accuracy of 69% for MI based BCI. Moreover, a 46% decline in IDR was observed when IC9(S1-S2) is compared with IC9(S1-S1), whereby IC9(S1-S1) achieved an accuracy of 90% for BCI competition IV-a dataset. Using

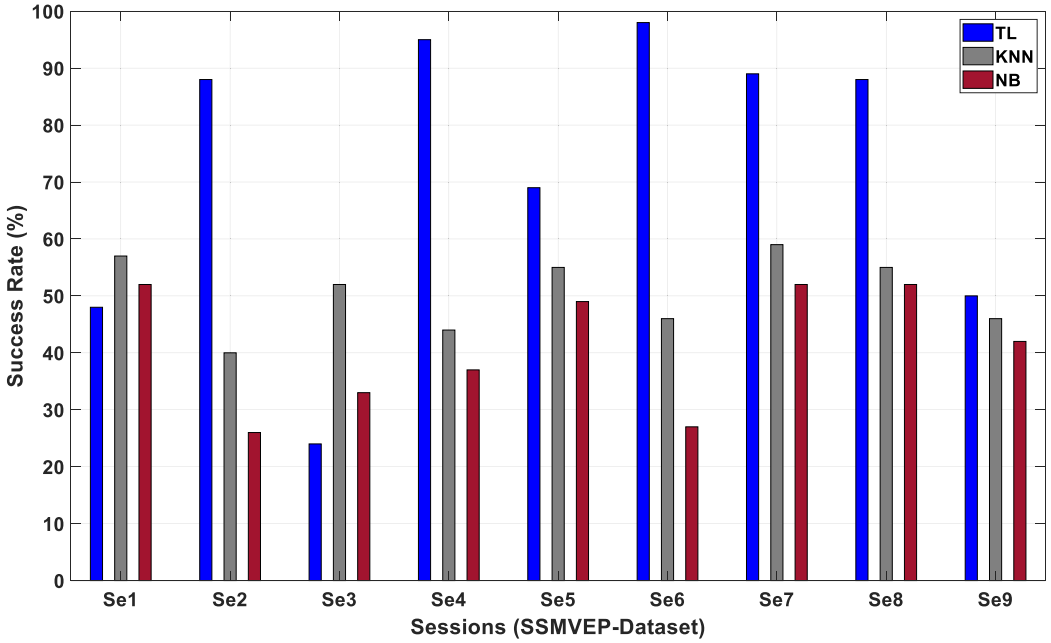


FIGURE 9. Classification results using nine SSMVEP sessions.

samples from a single subject for training, validation and testing the classifier yielded higher IDR results. However, a significant decline in IDRs were captured for all inter-subject experiments. If properly select ICs, the SSMVEP sessions can be more robust to the inter-subject variability, comparing to the MI sessions.

Conclusively, the inter-subject variability can significantly drop the IDR when features and classifiers are used across subjects. But there also exist some common characteristics among subjects, such as the IC12 in SSMVEP dataset demonstrate some clues of these common characteristics.

5) FACTOR 5: FEATURE CLASSIFICATION

In this section three classifiers (TL, K-NN, NB) are evaluated and compared, mainly to determine whether multi-source neural information can be used to improve CA of individual sessions, subjects or domains. As such when TL is implemented Se1 is treated as a target domain, while the remaining sessions (Se2∼Se9) are treated as source domains. However, when both K-NN and NB are implemented samples from (Se2∼Se9) are used to train the classifiers, while samples from Se1 are used to predict. In this case nine SSMVEP sessions acquired in different days are used to facilitate the investigation. Subsequently a highest accuracy of 98% was obtained for TL when Se6 is the target domain, while the rest of the sessions are source domains (Se1∼Se5 and Se7∼Se9) as illustrated in Figure 9. However, accuracies of 27% and 46% were observed for both K-NN and NB respectively, when samples from 8 sessions are used to train the classifiers and samples from a single session are used to test the classifier. In a similar manner TL achieved highest accuracies of 95% and 89% when Se4 and Se8 are target domains respectively. Moreover, accuracies of 37% and 52% were observed for NB, while 44% and 55% were observed for K-NN across Se4 and Se8. An accuracy of 88% was obtained for TL across Se2 and 69% across Se5 as target domains. However, a decline in accuracy was observed whereby K-NN achieved accuracies of 40% and 55%, while NB achieved accuracies of 26% and 49% across both Se2 and Se5 respectively. Furthermore, accuracies of 48% and 50% were obtained for TL when Se1 and Se9 are target domains, accuracies of 52% and 43% were observed for NB, while K-NN achieved accuracies of 57% and 46% respectively. Moreover, a lowest accuracy of 24% was obtained when Se3 is a target domain, while the rest of the sessions are source domains (Se1∼Se2 and Se4∼Se9). In this case noise, low concentration level or mental fatigue can affect transferability across domains, in turn constitute to low IDR when Se1, Se3 and Se9 are considered as target domain.

Based on this results it is worth noting that transfer learning approach can significantly improve classification accuracy, as compared to using samples from multiple sessions to train, while samples from a single session are used to test the classifier. In this case a significant decline in accuracy was observed for both K-NN and NB when samples from 8 different sessions are used to train and 1 session to test the classifier. However, a significant increase in accuracy was observed when knowledge from multiple source domains are used to enhance learning performance of the target domain as depicted in Figure 9.

Five SSMVEP subjects are further evaluated and compared using three classification methods (TL, K-NN, NB), mainly

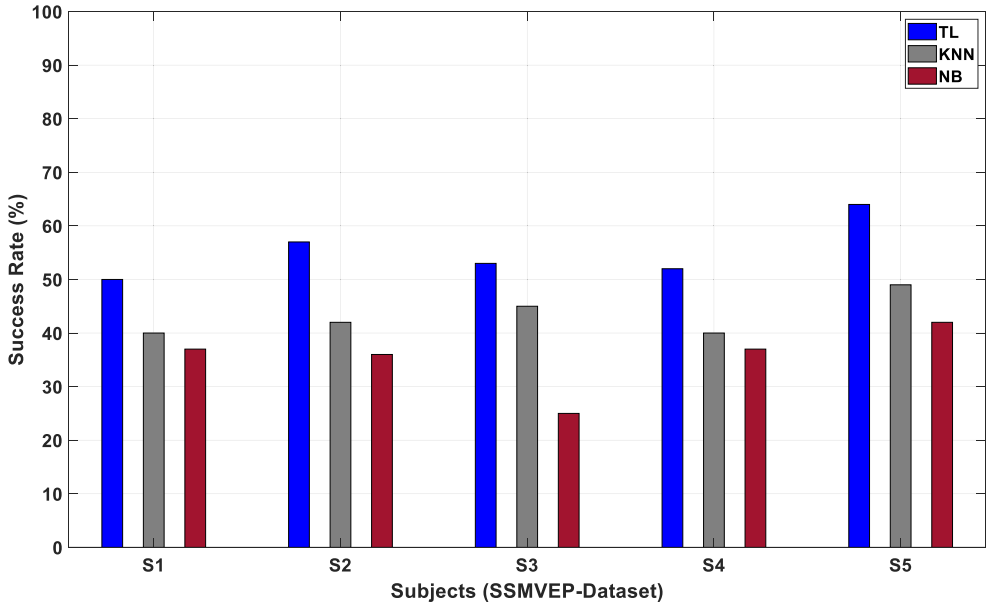


FIGURE 10. Classification results using five SSMVEP subjects.

TABLE 4. Average classification accuracy on three datasets.

Dataset	Classification Algorithms		
	TL	K-NN	NB
9 Sessions (SSMVEP-Dataset)	72%	50.4%	41.2%
5 Subjects (SSMVEP-Dataset)	56.8%	43%	35.4%
9 Subjects (BCI Competition IV-a)	76.6%	58%	42.4%

to determine whether multi-source neural information can be used to improve CA of individual sessions, subjects or domains. Subsequently when TL is implemented S1 is treated as a target domain, while the remaining subjects (S2∼S5) are treated as source domains. However, when both K-NN and NB are implemented samples from (S2∼S5) are used to train the classifiers, while samples from S1 are used to predict. In this case neural information from four subjects are considered as the source domains, while a single subject is treated as the target domain. Subsequently a highest accuracy of 64% was obtained when S5 is the target domain, while the rest of the subjects are source domains (S1∼S4) as illustrated in Figure 10. However, accuracies of 42% and 49% were observed for both K-NN and NB respectively, when samples from 4 subjects are used to train the classifiers and samples from a single subject are used to test the classifier. In a similar manner TL achieved an accuracy of 53% when both S3 and S4 are target domains respectively. Moreover, accuracies of 25% and 37% were observed for NB, while 45% and 40% were observed for K-NN across S3 and S4 respectively.

A lowest accuracy of 50% was achieved when S1 is the target domain and the other subjects (S2∼S5) are source domains. Subsequently, a decline in accuracy was observed when samples from S1 are used to test and samples from (S2∼S5) are used to train the classifier. As such both K-NN and NB achieved an accuracy of 40% and 37% respectively.

In this case evaluating commonalities across different subjects using samples from multiple subjects to train the classifier, and samples from another single subject to test the classifier resulted in a significant decline in accuracy as depicted in Figure 10. However, an increase in IDR was observed when TL is employed to transfer knowledge in the source domains to a target domain, while IDR deteriorated for both K-NN and NB when samples from five subjects are used to train, and samples from a single subject to test the classifiers. Moreover, based on this results it is also worth noting that all 5 subjects were BCI illiterates and had no prior BCI training before the experiment, as such a factor such as a feeling of tiredness was observed during SSMVEP signal acquisition, which may in turn affect the quality of EEG signal and resulting in low prediction rate.

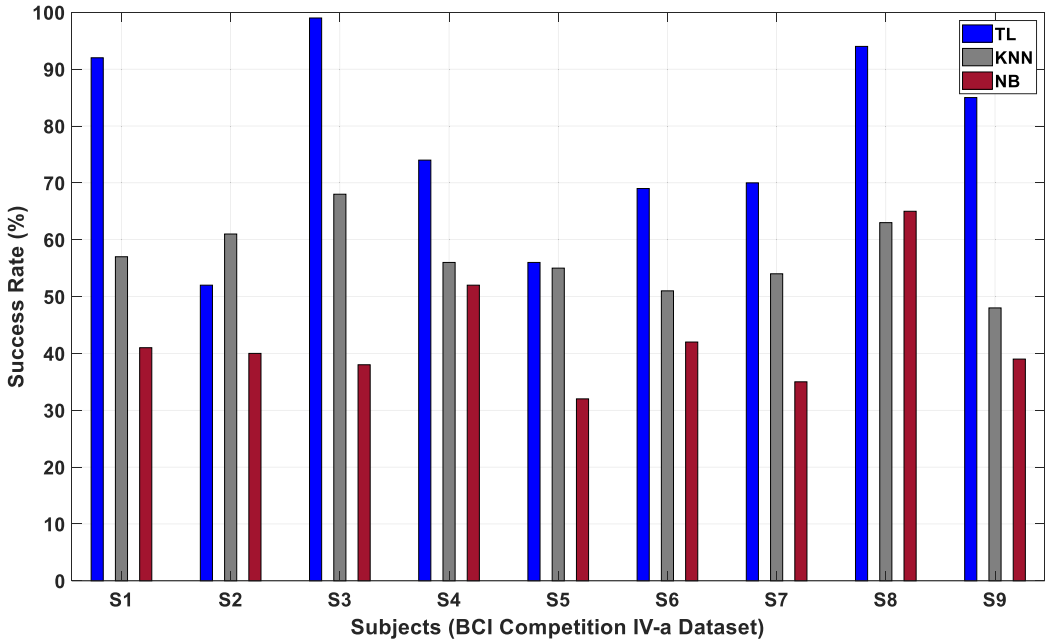


FIGURE 11. Classification results using nine MI subjects.

To further investigate whether neural information from various sources can improve IDR of individual sessions/subjects or domains, three classifiers (TL, K-NN, NB) are evaluated and compared in this section. BCI competition IV-a dataset (consisting of nine subjects performing MI tasks) is used to facilitate the investigation. As such S1 is treated as a target domain, while the remaining sessions (S2∼S9) are treated as source domains when TL is implemented. However, when both K-NN and NB are implemented samples from (S2∼S9) are used to train the classifiers, while samples from S1 are used to predict. Subsequently a highest accuracy of 99% was obtained for TL when S3 is the target domain, while the rest of the subjects are source domains (S1∼S2 and S4∼S9) as illustrated in Figure 11. However, accuracies of 68% and 38% were observed for both K-NN and NB respectively, when samples from 8 subjects are used to train the classifiers and samples from a single subject are used to test the classifier. In a similar manner TL achieved highest accuracies of 91% and 94% when S1 and S8 are target domains respectively. Moreover, accuracies of 41% and 65% were observed for NB, while 93% and 48% were observed for K-NN across S1 and S8. An accuracy of 70% was obtained for TL across S7 and 74% across S5 as target domains. However, a decline in accuracy was observed whereby NB achieved accuracies of 35% and 32%, while K-NN achieved accuracies of 54% and 55% across both S7 and S5 respectively. Furthermore, accuracies of 74% and 85% were obtained for TL when S4 and S9 are target domains, similarly accuracies of 52% and 39% were observed for NB, while K-NN achieved accuracies of 56% and 48% respectively. A further decline in accuracy was observed across S6, whereby NB achieved an accuracy of 42% and K-NN an accuracy of 51%, while an increase in accuracy was observed when S6 is the target domain, in this case TL achieved an accuracy of 69%. Moreover, a lowest accuracy of 52% was obtained when S2 is a target domain, while the rest of the sessions are source domains (S1 and S3∼S9).

From these results it is worth noting that learning tasks in unrelated domains during BCI training turns to show intense individual variations across subjects. However, a significant increase in accuracy or learning performance was observed across domains, when TL is employed to transfer features in the source domains to the target domains. In this case TL approach shows significant increase in classification accuracy as compared to both K-NN and NB when samples from 8 subjects are used to train, while samples from a different single subject are used to test the classifier. Moreover, significant variations in CA across all 9 target domains were observed, and based on these results it is also worth noting that variability in neural dynamics across different subjects can result in negative transfer, which in turn deteriorate CA across target domains.

The performance of all three classifiers is further evaluated to determine whether features from different sources can enhance prediction rate of individual sessions/subjects. TABLE 4 shows the average CA results, where TL firstly achieved a highest average accuracy of 72% as compared to both K-NN and NB, with highest average CAs of 50.4% and 41.2% respectively when 9 sessions are used for cross-session classification. Furthermore, when 5 subjects are considered for cross-subject classification, a highest average CA of 56.8% was observed when TL is employed, while lowest

average CA of 43% and 35.4% were achieved when both K-NN and NB are employed respectively as illustrated in TABLE 4. Moreover, TL also achieved a highest average accuracy of 76.6% as compared to both K-NN and NB, with average CAs of 58% and 42.4% respectively when 9 subjects are considered for cross-subject classification. In this case TL outperformed both K-NN and NB classifier, and based on this results it is worth noting that features from various sources can significantly enhance performance when TL is employed, while a significant decrease is observed when both NB and K-NN are employed.

V. DISCUSSION

BCI performance in the form of IDR from EEG signals is a significant component of EEG based BCIs. Although reducing signal-to-noise ratio of EEG signals through filtering and decomposition has proven to be important, other factors also demonstrate significant effects on the IDR. We picked few potential factors and studied their impacts on the IDR. The selection of ICs affects the performance of BCIs due to sessional diversity, as shown in the results in TABLE 1, where IDRs vary significantly among individual ICs, with the highest 100% IDR contrasting with the lowest 51% for SSMVEP, and 69% against 47% for MI tasks, for the same subject in the same session.

The alpha rhythm wave is recognized responsible for neural processes such as concentration, in this case the alpha rhythm magnitude validated the effects of changes in concentration level on IDR. The changes in alpha rhythm magnitudes were strongly associated to the changes of IDR, in an inverse proportional manner. A 13% increase in alpha rhythm magnitude was observed across Se2 and Se3 which resulted in a 13% decrease in accuracy for own MI dataset. As demonstrated in Figures 4, 5 and 6 in this paper, the results further validate that change in concentration level can be represented by an alpha rhythm wave. SSMVEP sessions seem more robust to the concentration changes, though the inverse proportional relationship exists as well.

Some known factors such as cognitive alteration, physiological and non-physiological factors can be tuned to minimize variability across sessions. However, the inter-session variability still significantly drops the IDRs across both intra and inter-sessions. A 21% and 15% decreases of IDR were observed across SSMVEP and MI sessions respectively.

Merging neural information from sessions acquired in different days resulted in a significant drop in intention detection rate. Though significant drops were observed, meaningful IDRs of 65% and 59% were achieved for SSMVEP and MI sessions when samples from Se1(IC5) were incorporated with samples from Se2(IC5), which means there are some common characteristics exist in different sessions, which can be used for inter-session tasks. In this case IC selection is still a crucial factor for inter-sessions IDRs.

A significant decline in IDR was observed when EEG signals acquired from different subjects were merged. The highest accuracy of 65% was observed for SSMVEP when samples from IC12 of the first subject (S1) were incorporated with samples from IC12 from the second subject (S2) for SSMVEP tasks. A maximum accuracy of 48% was observed for MI-tasks, when samples from IC11 of the first subject (S1) were incorporated with samples from IC11 of the second subject (S2). This is interpreted to significant inverse effects of inter-subject variability on IDRs, meanwhile, the existence of common inter-subject characteristics is also supported by the experimental results, although such characteristics might be weak.

Using neural information from multiple sources to enhance learning performance of the target domain resulted in a significant increase in CA, while a decline in CA was observed when samples from multiple sources are used to train the classifiers, and samples from a single source to predict. Transfer learning in this instance has yielded a significantly high accuracy as compared to both K-NN and NB classifiers using SSMVEP sessions, whereby a highest accuracy of 98% was achieved when Se6 is the target domain, and the remaining sessions (Se1∼Se5 and Se7∼Se8) are source domains as depicted in Figure 9. In this case both K-NN and NB achieved accuracies of 59% and 52% which is significantly low as compared to TL. In a similar manner TL yielded a highest accuracy of 64% as compared to both K-NN and NB using SSMVEP subjects, when S5 is the target domain and (S1∼S4) are source domains. K-NN yielded a highest accuracy of 49% when samples from (S1∼S4) are used to train and S5 to predict, while NB achieved a highest accuracy of 37% when samples from (S2∼S5) are used to train and S1 to predict as demonstrated in Figure 10. Furthermore, TL achieved a highest accuracy of 99% as compared to both K-NN and NB using BCI competition IV-a dataset, when S3 is the target domain and (S1∼S2 and S4∼S9) are source domains. K-NN obtained a highest accuracy of 68% when samples from (S1∼S2 and S4∼S9) are used to train and S3 to predict, while NB achieved a highest accuracy of 65% when samples from (S1∼S7 and S9) are used to train and S8 to predict as depicted in Figure 11.

VI. CONCLUSION

This study investigated the impact of five factors contributing to low IDR of EEG based BCIs, including the selection of ICs, changes in concentration level, inter-session and inter-subject variability, feature classification. Three datasets (BCI competition IV-a and our own recorded SSMVEP and MI datasets) were used to facilitate the investigation. After ICA, 16 independent components were obtained and used in all experiments. Significant varying IDRs were captured, corresponding to concentration levels indicated by the following factors: (i) alpha rhythm magnitude, (ii) variability within components, (iii) variability within sessions, (iv) variability within subjects, and (v) classification methods.

Experimental results supported the significance of the concerned factors on IDR of BCIs. Moreover, domain transfer learning approach was introduced, mainly to address the challenge of inter-session and inter-subject variability. In this case domain transfer learning resulted in a significant increase in accuracy across domains, while a significant decline in IDRs was observed when samples from multiple subjects/sessions are utilized to train the classifier and a single subject/session to test the classifiers (NB and K-NN) in a classical BCI.

Besides of the concerned factors, there are still more potential factors which might affect the IDR under various scenarios, such as the feature manipulation in spaces, domains, manifolds, etc. These factors will be considered in the future study.

REFERENCES

[1] H. Zeng and A. Song, "Optimizing single-trial EEG classification by stationary matrix logistic regression in brain–computer interface," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 11, pp. 2301–2313, Nov. 2016.

[2] X. Xie, Z. L. Yu, H. Lu, Z. Gu, and Y. Li, "Motor imagery classification based on bilinear sub-manifold learning of symmetric positive-definite matrices," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 504–516, Jun. 2017.

[3] Y. Chu, X. Zhao, Y. Zou, W. Xu, and Y. Zhao, "Robot-assisted rehabilitation system based on SSVEP brain–computer interface for upper extremity," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2018, pp. 1098–1103.

[4] C. Y. Sai, N. Mokhtar, H. Arof, P. Cumming, and M. Iwahashi, "Automated classification and removal of EEG artifacts with SVM and wavelet-ICA," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 3, pp. 664–670, May 2018.

[5] R. Masoomi and A. Khadem, "Enhancing LDA-based discrimination of left and right hand motor imagery: Outperforming the winner of BCI competition II," in *Proc. 2nd Int. Conf. Knowl.-Based Eng. Innov. (KBEI)*, Nov. 2015, pp. 392–398.

[6] W.-Y. Hsu, C.-H. Lin, H.-J. Hsu, P.-H. Chen, and I.-R. Chen, "Wavelet-based envelope features with automatic EOG artifact removal: Application to single-trial EEG data," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 2743–2749, 2012.

[7] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 121–131, Jan. 2011.

[8] A. Sharmila and P. Geethanjali, "DWT based detection of epileptic seizure from EEG signals using Naive Bayes and k-NN classifiers," *IEEE Access*, vol. 4, pp. 7716–7727, 2016.

[9] E. Alickovic, J. Kevric, and A. Subasi, "Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction," *Biomed. Signal Process. Control*, vol. 39, pp. 94–102, Jan. 2018.

[10] E. Alickovic and A. Subasi, "Ensemble SVM method for automatic sleep stage classification," *IEEE Instrum. Meas.*, vol. 67, no. 6, pp. 1258–1265, Jun. 2018.

[11] T. K. Reddy, Y.-K. Wang, C.-T. Lin, and J. Andreu-Perez, "Joint approximate diagonalization divergence based scheme for EEG drowsiness detection brain computer interfaces," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jul. 2021, pp. 1–6.

[12] T. K. Reddy, V. Arora, L. Behera, Y.-K. Wang, and C.-T. Lin, "Fuzzy divergence based analysis for EEG drowsiness detection brain computer interfaces," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jul. 2020, pp. 1–7.

[13] S. Kumar, T. K. Reddy, V. Arora, and L. Behera, "Formulating divergence framework for multiclass motor imagery EEG brain computer interface," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1344–1348.

[14] S. Saha, K. I. U. Ahmed, R. Mostafa, L. Hadjileontiadis, and A. Khandoker, "Evidence of variabilities in EEG dynamics during motor imagery-based multiclass brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 371–382, Feb. 2018.

[15] S. Saha, K. I. Ahmed, R. Mostafa, A. H. Khandoker, and L. Hadjileontiadis, "Enhanced inter-subject brain computer interface with associative sensorimotor oscillations," *Healthcare Technol. Lett.*, vol. 4, no. 1, pp. 39–43, Feb. 2017.

[16] M.-H. Lee, O.-Y. Kwon, Y.-J. Kim, H.-K. Kim, Y.-E. Lee, J. Williamson, S. Fazli, and S.-W. Lee, "EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy," *GigaScience*, vol. 8, no. 5, May 2019, Art. no. giz002.

[17] R. M. Mehmood, R. Du, and H. J. Lee, "Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors," *IEEE Access*, vol. 5, pp. 14797–14806, 2017.

[18] V. S. Handiru and V. A. Prasad, "Optimized bi-objective EEG channel selection and cross-subject generalization with brain–computer interfaces," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 6, pp. 777–786, Dec. 2016.

[19] B. A. Osuagwu, M. Zych, and A. Vuckovic, "Is implicit motor imagery a reliable strategy for a brain–computer interface?" *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 12, pp. 2239–2248, Dec. 2017.

[20] M. Nakanishi, Y. Wang, X. Chen, Y. Wang, X. Gao, and T.-P. Jung, "Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 1, pp. 104–112, Jan. 2018.

[21] S. Liu, J. Tong, M. Xu, J. Yang, H. Qi, and D. Ming, "Improve the generalization of emotional classifiers across time by using training samples from different days," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 841–844.

[22] S. Liu, J. Meng, J. Yang, X. Zhao, F. He, H. Qi, and D. Ming, "Within-stimulus emotion recognition may inflate the classification accuracies based on EEG signals," in *Proc. IEEE 7th Int. Conf. Awareness Sci. Technol. (iCAST)*, Sep. 2015, pp. 115–118.

[23] C.-S. Wei, Y.-T. Wang, C.-T. Lin, and T.-P. Jung, "Toward drowsiness detection using non-hair-bearing EEG-based brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 400–406, Feb. 2018.

[24] C. Maswanganyi, C. Tu, P. Owolawi, and S. Du, "Discrimination of motor imagery task using wavelet based EEG signal features," in *Proc. Int. Conf. Intell. Innov. Comput. Appl. (ICONIC)*, Dec. 2018, pp. 1–4.

[25] C. Maswanganyi, C. Tu, P. Owolawi, and S. Du, "Overview of artifacts detection and elimination methods for BCI using EEG," in *Proc. IEEE 3rd Int. Conf. Image, Vis. Comput. (ICIVC)*, Jun. 2018, pp. 832–836.

[26] D. Rathee, H. Raza, G. Prasad, and H. Cecotti, "Current source density estimation enhances the performance of motor-imagery-related brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 12, pp. 2461–2471, Dec. 2017.

[27] A. Bhardwaj, A. Gupta, P. Jain, A. Rani, and J. Yadav, "Classification of human emotions from EEG signals using SVM and LDA classifiers," in *Proc. 2nd Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Feb. 2015, pp. 180–185.

[28] S. Siuly and Y. Li, "Improving the separability of motor imagery EEG signals using a cross correlation-based least square support vector machine for brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 4, pp. 526–538, Jul. 2012.

[29] L. F. Nicolas-Alonso, R. Corralejo, J. Gomez-Pilar, D. Álvarez, and R. Hornero, "Adaptive stacked generalization for multiclass motor imagery-based brain computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 4, pp. 702–712, Jul. 2015.

[30] T. Yu, J. Xiao, F. Wang, R. Zhang, Z. Gu, A. Cichocki, and Y. Li, "Enhanced motor imagery training using a hybrid BCI with feedback," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 7, pp. 1706–1717, Jul. 2015.

[31] Y. Wang, X. Chen, X. Gao, and S. Gao, "A benchmark dataset for SSVEP-based brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.* vol. 25, no. 10, pp. 1746–1752, Nov. 2016.

[32] H. Ghandeharion and A. Erfanian, "A fully automatic ocular artifact suppression from EEG data using higher order statistics: Improved performance by wavelet analysis," *Med. Eng. Phys.*, vol. 32, pp. 720–729, Sep. 2010.

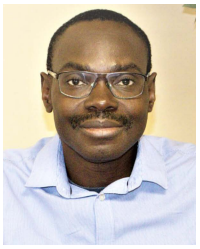
[33] F. P. A. Lestari, E. S. Pane, Y. K. Suprapto, and M. H. Purnomo, "Wavelet based-analysis of alpha rhythm on EEG signal," in *Proc. Int. Conf. Inf. Commun. Technol. (ICOIACT)*, Mar. 2018, pp. 719–723.

[34] L. Bi, X.-A. Fan, and Y. Liu, "EEG-based brain-controlled mobile robots: A survey," *IEEE Trans. Human-Mach. Syst.*, vol. 43, no. 2, pp. 161–176, Mar. 2013.

[35] E. F. González-Castañeda, A. A. Torres-García, C. A. Reyes-García, and L. Villaseñor-Pineda, "Sonification and textification: Proposing methods for classifying unspoken words from EEG signals," *Biomed. Signal Process. Control*, vol. 37, pp. 82–91, Aug. 2017.
[36] L. Qin, B. Kamousi, Z. Liu, L. Ding, and B. He, "Classification of motor imagery tasks by means of time-frequency-spatial analysis for brain–computer interface applications," in *Proc. 2nd Int. IEEE EMBS Conf. Neural Eng.*, Mar. 2005, pp. 374–376.
[37] V. Bono, S. Das, W. Jamal, and K. Maharatna, "Hybrid wavelet and EMD/ICA approach for artifact suppression in pervasive EEG," *J. Neurosci. Methods*, vol. 267, pp. 89–107, Jul. 2016.
[38] A. Subasi and M. I. Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8659–8666, 2010.
[39] J. Zhuang and G. Yin, "Motion control of a four-wheel-independent-drive electric vehicle by motor imagery EEG based BCI system," in *Proc. 36th Chin. Control Conf. (CCC)*, Jul. 2017, pp. 5449–5454.
[40] P. K. Pattnaik and J. Sarraf, "Brain computer interface issues on hand movement," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 30, no. 1, pp. 18–24, Jan. 2018.
[41] Y. U. Khan and F. Sepulveda, "Brain–computer interface for single-trial EEG classification for wrist movement imagery using spatial filtering in the gamma band," *IET Signal Process.*, vol. 4, no. 5, p. 510, 2010.
[42] H. Zhang, H. Yang, and C. Guan, "Bayesian learning for spatial filtering in an EEG-based brain–computer interface," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 7, pp. 1049–1060, Jul. 2013.
[43] S. Z. Bong, K. Wan, M. Murugappan, N. M. Ibrahim, Y. Rajamanickam, and K. Mohamad, "Implementation of wavelet packet transform and non linear analysis for emotion classification in stroke patient using brain signals," *Biomed. Signal Process. Control*, vol. 36, pp. 102–112, Jul. 2017.
[44] T. Mladenov, K. Kim, and S. Nooshabadi, "Accurate motor imagery based dry electrode brain-computer interface system for consumer applications," in *Proc. IEEE 16th Int. Symp. Consum. Electron.*, Jun. 2012, pp. 1–4.
[45] N. A. Abduallatif, S. G. Elsherbini, B. S. Boshra, and I. A. Yassine, "Brain–computer interface controlled functional electrical stimulation system for paralyzed arm," in *Proc. 8th Cairo Int. Biomed. Eng. Conf. (CIBEC)*, Dec. 2016, pp. 48–51.
[46] S. O'Regan, S. Faul, and W. Marnane, "Automatic detection of EEG artefacts arising from head movements using EEG and gyroscope signals," *Med. Eng. Phys.*, vol. 35, pp. 867–874, Jul. 2013.
[47] R. N. Khushaba, A. Al-Ani, and A. Al-Jumaily, "Feature subset selection using differential evolution and a statistical repair mechanism," *Expert Syst. Appl.*, vol. 38, no. 9, pp. 11515–11526, 2011.
[48] R. Lahiri, P. Rakshit, and A. Konar, "Evolutionary perspective for optimal selection of EEG electrodes and features," *Biomed. Signal Process. Control*, vol. 36, pp. 113–137, Jul. 2017.
[49] L. Xie, Z. Deng, P. Xu, K.-S. Choi, and S. Wang, "Generalized hidden-mapping transductive transfer learning for recognition of epileptic electroencephalogram signals," *IEEE Trans. Cybern.*, vol. 49, no. 6, pp. 2200–2214, Jun. 2019.
[50] M. Dai, S. Wang, D. Zheng, R. Na, and S. Zhang, "Domain transfer multiple kernel boosting for classification of EEG motor imagery signals," *IEEE Access*, vol. 7, pp. 49951–49960, 2019.
[51] M.-A. Li and D.-Q. Xu, "A transfer learning method based on VGG-16 convolutional neural network for MI classification," in *Proc. 33rd Chin. Control Decis. Conf. (CCDC)*, May 2021, pp. 5430–5435.
[52] D. Ming, Y. Xi, M. Zhang, H. Qi, L. Cheng, B. Wan, and L. Li, "Electroencephalograph (EEG) signal processing method of motor imaginary potential for attention level classification," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2009, pp. 4347–4351.
[53] A. Myrden and T. Chau, "A passive EEG-BCI for single-trial detection of changes in mental state," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 4, pp. 345–356, Apr. 2017.
[54] J. J. Bird, J. Kobylarz, D. R. Faria, A. Ekárt, and E. P. Ribeiro, "Cross-domain MLP and CNN transfer learning for biological signal processing: EEG and EMG," *IEEE Access*, vol. 8, pp. 54789–54801, 2020.
[55] Y. Qin, N. Zhang, Y. Chen, Y. Tan, L. Dong, P. Xu, D. Guo, T. Zhang, D. Yao, and C. Luo, "How alpha rhythm spatiotemporally acts upon the thalamus-default mode circuit in idiopathic generalized epilepsy," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 4, pp. 1282–1292, Apr. 2021.
[56] L. Zhang, X. Wu, X. Guo, J. Liu, and B. Zhou, "Design and implementation of an asynchronous BCI system with alpha rhythm and SSVEP," *IEEE Access*, vol. 7, pp. 146123–146143, 2019.

RITO CLIFFORD MASWANGANYI received the master's degree in computer systems engineering from the Tshwane University of Technology, South Africa, in 2019, where he is currently pursuing the Ph.D. degree in computer systems engineering with the Faculty of Information and Communication Technology. His research interests include signal processing, pattern recognition, and brain–computer interface.

CHUNLING TU (Member, IEEE) received the bachelor's degree in computer science from the Tianjin University of Technology and Education, China, in 2002, the M.Tech. degree in electrical engineering from the Tshwane University of Technology, South Africa, the M.Sc. degree in electrical engineering from ESIEE Paris University, France, in 2010, the D.Tech. degree in electrical engineering from the Tshwane University of Technology, and the Ph.D. degree in electrical engineering from University Paris East, France, in 2015. She is currently an Associate Professor at the Tshwane University of Technology. Her research interests include image processing, AI, industrial control, machine learning, deep learning, and pattern recognition.



PIUS ADEWALE OWOLAWI received the bachelor's degree from the Federal University of Technology, Akure, Nigeria, in 2001, and the master's and Ph.D. degrees in electrical engineering from the University of Kwazulu Natal, South Africa, in 2006 and 2010, respectively. He is currently the Head of the Department of Computer Systems Engineering, Tshwane University of Technology, South Africa. His research interests include RF, green communication, radiowave propagation (microwave/ millimeter wave systems), satellite and free space optical communications, the IoT, embedded systems, machine learning, and data analytics. He was a recipient of the Joint Holder of Best Paper Award for a paper presented at the 2nd International Conference on Applied and Theoretical Information Systems Research, Taipei, Taiwan, in 2012, and the Vice Chancellor's Teaching Excellence Award in 2015.

SHENGZHI DU received the M.S. degree in control theory and control engineering from Tianjin Polytechnic University, Tianjin, China, in 2001, and the Ph.D. degree in control theory and control engineering from Nankai University, Tianjin, in 2005. He is currently a Full Professor with the French South Africa Institute of Technology (F'SATI) and the Tshwane University of Technology, South Africa. His research interests include computer vision, AI, pattern recognition, and human in the loop systems.