Transfer learning of an ensemble of DNNs for SSVEP BCI spellers without user-specific training

Osman Berke Guney1,∗ and Huseyin Ozkan2

1 Department of Electrical and Computer Engineering, Boston University, Boston, MA, United States of America
2 Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey
∗ Author to whom any correspondence should be addressed.
E-mail: berke@bu.edu

Keywords: steady state visually evoked potential, brain computer interfaces, ensemble, deep learning, transfer learning.

Abstract
Objective. Steady-state visually evoked potentials (SSVEPs), measured with electroencephalogram (EEG), yield decent information transfer rates (ITRs) in brain-computer interface (BCI) spellers. However, the current high performing SSVEP BCI spellers in the literature require an initial lengthy and tiring user-specific training for each new user for system adaptation, including data collection with EEG experiments, algorithm training and calibration (all are before the actual use of the system). This impedes the widespread use of BCIs. To ensure practicality, we propose a novel target identification method based on an ensemble of deep neural networks (DNNs), which does not require any sort of user-specific training. Approach. We exploit already-existing literature datasets from participants of previously conducted EEG experiments to train a global target identifier DNN first, which is then fine-tuned to each participant. We transfer this ensemble of fine-tuned DNNs to the new user instance, determine the $k$ most representative DNNs according to the participants’ statistical similarities to the new user, and predict the target character through a weighted combination of the ensemble predictions. Main results. The proposed method significantly outperforms all the state-of-the-art alternatives for all stimulation durations in [0.2–1.0] s on two large-scale benchmark and BETA datasets, and achieves impressive 155.51 bits/min and 114.64 bits/min ITRs. Code is available for reproducibility: https://github.com/osmanberke/Ensemble-of-DNNs. Significance. Our Ensemble-DNN method has the potential to promote the practical widespread deployment of BCI spellers in daily lives as we provide the highest performance while enabling the immediate system use without any user-specific training.

1. Introduction

Brain-computer interfaces (BCIs) enable motor disabled individuals to communicate with [1, 2] and control their surroundings [3, 4]. BCIs can also be used for stroke rehabilitation [5, 6]. Additionally, novel stimulation paradigms [7] as well as advanced signal processing and machine learning techniques [8] have been extensively studied to improve BCIs. A prominent research direction focuses on speller designs in which the user can type alphanumeric characters through brain signals alone by using a visual interface [9]. Electroencephalogram (EEG) is widely used in BCI applications to measure the brain activity non-invasively, providing a direct transmission pathway between a computer and the brain [10]. On the other hand, thanks to its high signal-to-noise ratio (SNR), reduced training time and high information transfer rate (ITR), steady state visually evoked potential (SSVEP) has attracted attention in the field recently and is now one of the leading paradigms [9, 10]. SSVEP, measured by EEG, is the brain’s response to a visual stimulus calibrated at a constant frequency. The spectrum of the resulting EEG waveform includes strong components that are tuned to the frequency of the input stimulus and its harmonics [11].

In SSVEP BCI spellers, the target character identification relies on frequency tagging that refers to the assignment of distinct oscillation frequencies to the
visuals of multiple alphanumeric characters. A person sees a matrix of flickering alphanumeric characters (i.e. uniquely tagged with frequencies) on the computer screen. S/he is asked to select and focus on the target character that is to be spelled at that moment. The objective is to correctly predict the target character based on the received SSVEP EEG signal.

One of the main concerns of the SSVEP BCI speller studies is to ease the system design for convenient implementations in real life. It is naturally not practical to collect data from each new-coming user (with lengthy and tiring preparation or calibration procedures or separate EEG experiments, all for algorithm training and system adaptation) before s/he actually starts using the system. We call these additional processes altogether user-specific training throughout the paper. Therefore, methods that do not require user-specific training are certainly preferable [12]. Nevertheless, one typically has some existing SSVEP EEG speller data prior to a new user arrives. This data might be from previous users, or persons participated in previous experiments or from the literature benchmarks.

Our goal in this paper is to exploit this pre-existing data by using transfer and ensemble learning ideas so that a new user can start using the BCI system immediately (no user-specific training) with a high enough performance thanks to the transferred knowledge. For this purpose, we propose a novel target identification method for SSVEP BCI speller systems (cf figure 1), which does not require user-specific training for a new user. Hence, as the main contribution of our study, we remove the hassle of long and tiring user-specific training periods present in the existing high performance spellers such as [11, 13]. Thus, the new user bears no hassle and her/his comfort is maximized.

Our method (cf figure 1) uses transfer and ensemble learning approaches in the sense that models that have been previously trained for different participants provide the ensemble and are transferred to the new user. The target character is then identified based on a weighted combination of the predictions of $k$ most representative models (i.e. the models of $k$ many participants whose data are statistically the most similar to that of the new user) from the ensemble. By using a dynamic selection, our method determines the best $k$ on the fly in the course of the target identification. Hence, we do not require any separate parameter tuning or parameter optimization or cross-validation steps. To construct the ensemble, we utilize a deep neural network (DNN) architecture that has been (figure 1) recently introduced in [11]. On two publicly available and widely used datasets (the benchmark [14] and BETA [15] datasets from 35 and 70 participants with 40 target

Figure 1. Previously collected data of $N$ different participants are used to train a global DNN. © [2022] IEEE. Adapted, with permission, from [11]. This global DNN is fine-tuned to each participant which yields $N$ participant-specific fine-tuned DNNs. In the testing phase, the most representative $k$ fine-tuned DNNs are selected according to the statistical similarities between the new user and participants. Predictions of these $k$ DNNs are combined in a weighted manner to make the final target character prediction (target identification) for the new user.
characters), we observe impressive 155.51 bits/min (on the benchmark) and 114.36 bits/min (on BETA) ITR performance figures. With these results, to the best of our knowledge, our proposed method is the best performing one among the most prominent literature methods that do not require user-specific training.

The paper organization is as follows. We present the related work in section 2. Our problem description and method are in sections 3 and 4. We report our performance evaluations in section 5 and conclude in section 6.

2. Related work

The SSVEP BCI speller systems in the literature can be categorized into two groups. Although the methods in the first group (e.g. [11, 13, 16]) typically yield better target identification accuracy, they (unlike ours) require lengthy and tiring user-specific training. Hence, this first group of methods are not comparable to our proposed method here. In our preliminary short conference proceeding (in Turkish) [16], we used the same ensemble of fine-tuned DNNs but combined all of them with weights given by a variant of the AdaBoost algorithm. In contrast, in this presented study, we combined only k most representative fine-tuned DNNs based on a weighting devised through a novel correlation based similarity measure and dynamic selection, which is fundamentally different. Moreover, [16] requires user-specific training thus belongs to the first group, whereas the presented study does not.

Like ours, the methods in the second group do not require any user-specific training. Either they are (a) completely training-free, or they rely on (b) transferring knowledge from previous models/training participants as prior information. These methods allow the new user to comfortably start using the system immediately with no hassle. We explain this second group of methods below in three paragraphs.

In the PSDA method [17], the target character is decoded as the one corresponding to the frequency of the highest SNR signal from all possible stimulation frequencies. In the MEC method [18], the received EEG signals are linearly combined to cancel the noise components and increase the SNR. On the other hand, the canonical correlation analysis (CCA) method [19] determines the correlation between the SSVEP signals and synthetically generated reference signals (composition of sinusoids of the stimulation frequencies) by finding the channel and harmonic combinations that maximize the correlation. The frequency that gives the maximum correlation is chosen as the target. Previous research has shown that the CCA method performs better than PSDA and MEC in terms of both the accuracy and ITR [19, 20]. As an improved extension of CCA, a method based on a filter-bank (multiple band-pass filters) approach with canonical correlation analysis (FBCCA) is proposed in [9]. These band-pass filters are used to extract the sub-band components from the received EEG signals and then CCA is applied to each sub-band separately [9]. The target character is predicted by combining the results of CCA.

In the tt-CCA method [21], the target character is predicted based on the correlation between the received test signal and certain template signals. For each character, all of the existing participants’ data (having the same label with the character in hand) are averaged to form a template signal. Then, CCA is applied between this template signal and the synthetic reference signal (composition of sinusoids of the corresponding frequency and its harmonics) so that the channel combination of the maximum correlation is selected. After completing this for all characters, the template/reference signals and channel combinations are transferred to the new user. While predicting the target character, tt-CCA combines three correlation coefficients for each character, and the one of the maximum combined correlation coefficient is the prediction. The first correlation coefficient is the maximum CCA coefficient between the new user instance and the reference signal. The channel combination of this maximum CCA coefficient combines the channels of the new user instance and the template signal, and then the correlation coefficient between them is calculated as the second. For the third coefficient, the same procedure is applied but with the transferred channel combination. Combined-tCCA [22] follows a similar approach, except that it uses different correlation coefficients defined in [23]. Rather than employing user-specific coefficients, those of [23] are modified in [22] by only using transferred signals and channel combinations. In the ttf-CCA method [24], the channel combinations are learned for each participant separately. After reducing the participant-specific channel combinations to certain common combinations, the correlation coefficients for each character between the new user instance and the corresponding template signal (similar to the templates of the tt-CCA method) are calculated. Maximum of these coefficients and another coefficient between the new user instance and the related synthetic template reveals the target prediction.

There are also certain other transfer learning based studies in the literature of SSVEP signal detection [25–27] and general EEG BCI applications [28, 29]. We refer the interested reader to [30] for a detailed literature review of transfer learning in this field. Each of the studies in [25–27] proposes a different neural network architecture for the same goal of generalizing well to the unseen data of a new user. For example, a variation of the architecture EEGNet [28] (EEGNet is originally for processing the EEG signals
First, we construct an ensemble of target identification. Then, we identify the most similar participants’ models at the target prediction stage for the new user. Hence, our proposed Ensemble-DNN method trains an ensemble of various models based on participants’ data and uses the predictions of most similar participants’ models at the target prediction stage for the new user. On the contrary, our proposed Ensemble-DNN method trains an ensemble of various models based on participants’ data and uses the predictions of most similar participants’ models at the target prediction stage for the new user. Consequently, we explicitly take into account the statistical variations across persons. On the contrary, our proposed Ensemble-DNN method trains an ensemble of various models based on participants’ data and uses the predictions of most similar participants’ models at the target prediction stage for the new user. Hence, our proposed Ensemble-DNN is significantly different and novel as a transfer learning based SSVEP signal detection method.

3. Problem description

We consider an $M$-character SSVEP BCI speller target identification ($M$-class classification) task for a multidimensional brain response (from a new user) $\mathbf{x} = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\} \in \mathbb{R}^{C \times N_1 \times N_2}$, measured by EEG using $C$ channels/electrodes with $N_1$ sample points (i.e., $N_1 = T \times F_s$ is the number of time samples with $F_s$ being the sampling frequency in Hz and $T$ being the signal length in seconds). Three band-pass filters ($N_2 = 3$) are used to extract the sub-bands of harmonics with no other preprocessing. Each of these filters has a different lower cut-off frequency $8 \times s$ (1 $\leq s \leq N_2$ is the filter index) and has the same upper cut-off frequency (90 Hz). We call this SSVEP EEG signal $\mathbf{x}$ (to be target-identified) that is received from the new user during a spelling session as the new user SSVEP EEG speller instance or simply the new user instance.

Our goal is to propose a novel SSVEP BCI speller target identification method (cf figure 1) that is based on training an ensemble of DNNs with previously existing data from various participants. This ensemble is directly transferred to the new user with no user-specific training (i.e., with no lengthy and tiring preparation or calibration procedures or separate EEG experiments, all for algorithm training and system adaptation), and then a weighted combination of the target character predictions of a certain subset of the ensemble DNNs is used for making the final target identification, i.e., final target character prediction/classification for the new user. In addition to the target identification accuracy, we also use the (ITR [31]) $\text{ITR} = (\log_2 M + P \log_2 P + (1 - P) \log_2 \left( \frac{1 - P}{M - 1} \right) \frac{C}{I_T}$ to measure the performance of our proposed method, where $P$ is the target identification accuracy. Note that the ITR formula is valid under certain assumptions [31]. For example, the probabilities of selecting non-target characters are assumed to be same ($\frac{P}{M}$) [31]. Although this assumption does not necessarily hold in practice, which can be seen from the nonuniform confusion matrices in figures 5 and 7 of [11], ITR is still a commonly used performance metric in the SSVEP BCI literature.

**Ensemble constituent DNN:** We use the DNN architecture of [11] (also shown in figure 1) to constitute our ensemble. This DNN is an end-to-end system that consists of four convolutional layers and a single fully connected layer. The first convolutional layer is for sub-band combination while the second convolutional layer is for channel combinations. The third and fourth convolutional layers are used to extract features after applying filters. Finally, the fully connected layer predicts the target character by choosing the character of the highest probability returned by the softmax applied in the end. We use three sub-bands and nine channels (Pz, PO3, PO5, PO4, PO6, POz, Oz1, Oz2, O2), whose physical locations on the skull are indicated in figure 1.

4. Method

(a) First, we construct an ensemble of target identification DNNs $\{f_{w(n)}^{(1)}\}_{n=1}^{N}$ of size $N$, each of which is trained based on the labeled $M$-class SSVEP EEG speller data from a (previously existing) different participant $n$. Overall, we have $N$ participants, $1 \leq n \leq N$. Here, $w(n)$ parameterizes the $n$th DNN with $f_{w(n)}(\mathbf{x}) = \hat{y}_n \in \{1, 2, \ldots, M\}$ being the target character prediction for the SSVEP EEG signal $\mathbf{x}$ from the new user.

(b) Then, we identify the $k$ most representative DNNs $\{f_{w(1)}^{(1)}, f_{w(2)}^{(1)}, \ldots, f_{w(k)}^{(1)}\}$ (here, $I(\cdot)$ is the indexing) after sorting the DNNs with respect to decreasing similarities between the new user’s data and participants’ data ($f_{w(1)}^{(1)}$ is the most representative DNN or the data of the $I(1)$th participant is statistically the most similar to that of the new user). We introduce a novel similarity measure $\rho_n$ for this purpose.

(c) Finally, we obtain the final target identification for the SSVEP EEG signal $\mathbf{x}$ from the new user.
as a weighted linear combination of the $k$ most representative DNN predictions as \( f(x) = \hat{y} = \arg \max_{i \in \{1, \ldots, M\}} \sum_{j=1}^{k} \nu_{i,j} \cdot f_{i}(j) \cdot I_{j}(i) \). Here, \( I_{j}(i) \) is the indicator function which outputs 1 if the condition is satisfied (otherwise 0) and \( \nu_{i,j} \)’s are the combination weights that use our similarity measure. We also propose a dynamic selection algorithm for determining an appropriate value for the parameter $k$. Hence, as an additional merit that further improve the new user comfort, our technique does not require any cross-validation or parameter optimization for $k$ as we select it on the fly as a part of our final target character prediction.

Remark 1. Our method does not require user-specific training for two reasons: (a) Our ensemble of DNNs can be trained and constructed offline even before a new user arrives, and (b) all of our computations use only the new user instance (besides participants’ data). Note that the new user instance is the SSVEP EEG signal that is received online during a spelling session and is the one to be target-identified. Hence, our method can be run in the online setting without requiring any offline data collection, so any new-coming user can immediately start using our BCI system with no hassle.

Remark 2. We point out that our method does successfully take into account two sources of statistical variations in the EEG signals: (a) The person-to-person statistical variations by using only the $k$ most representative participant-specific DNNs (based on participant-to-new user statistical similarities) rather than using all of them, and (b) the within-person statistical variations by repeating the dynamic selection of the parameter $k$ for every new user instance.

Next, we expand on the details below in the same order.

4.1. Training of the ensemble \( \{f_{w(a)}\}_{a=1}^{N} \) of DNNs

Our training strategy for obtaining the ensemble of DNNs is similar to the procedure in [11] in one aspect since we also use two-stage training, but completely different in another aspect since we do not use user-specific training. (a) In the first stage, a single global DNN is trained based on the entire set of labeled $M$-class SSVEP EEG speller data from pre-existing $N$ participants. (b) In the second stage, the global DNN is fine-tuned to each of the participants based only on her/his own data. Thus, an ensemble \( \{f_{w(a)}\}_{a=1}^{N} \) of fine-tuned DNNs is obtained. The first layer weights of the DNN architecture are initialized with unity, and the initialization of other layers’ weights are sampled from Gaussian distribution with 0 mean and standard deviation 0.01. Also, 0.1, 0.1 and 0.95 dropout probabilities are applied between the second and third, third and fourth, and fourth and fifth layers, respectively. The network is trained by minimizing the categorical cross entropy loss \( \frac{1}{N} \sum_{a=1}^{N} -\log(s_{a,q_{a}}) + \lambda \|w\|^{2} \), where $\lambda = 0.001$ is the weight of the L2 regularization, $D_{b}$ is the number of SSVEP EEG speller instances in the batch, $s_{a,q_{a}}$ is the $q_{a}$th index of the softmax output for the corresponding instance with the true label $q_{a}$, and $w$ is the weights of the all layers in the DNN. This network setup and parameter initializations are same as the ones in [11]. On the contrary, we emphasize that our target identification in this paper is an ensemble approach and is free of user-specific training/data as described, which is fundamentally different from [11] where the target character prediction is made by a single network in a supervised manner that strictly requires labeled user-specific data.

4.2. The most representative $k$ participants \( \{f_{w(I(j))}\}_{j=1}^{k} \)

We use each participant-specific fine-tuned DNN in the ensemble \( \{f_{w(a)}\}_{a=1}^{N} \) to produce a target character prediction for an SSVEP EEG speller instance $x$ of the new user, yielding an ensemble \( \{f_{w(a)}(x) = \hat{y}_{a}\}_{a=1}^{N} \) of predictions. However, since the EEG signal statistics are well known to significantly vary from one person to another, only a $k$-subset of our introduced ensemble is expected to be relevant to the final target identification. Namely, the DNNs of the participants that are more similar to the new user should provide more reliable predictions. Therefore, we devise a subset selection process for determining the most representative $k$ participants whose data are statistically the most similar (correlative) to that of the new user, which provides a $k$-subset of the DNN ensemble as \( \{f_{w(I(j))}\}_{j=1}^{k} \) with $I(\cdot)$ being the indexing (descending similarities). For this purpose, below we introduce a novel similarity measure between the participants and the new user.

Our similarity measure is based on the new user’s SSVEP EEG speller instance, as well as the participants’ instances that are correctly labeled by their respective fine-tuned DNNs. We start with defining an individual template \( \mathbf{x}_{n}^{w} \), which is the mean of the $n$th fine-tuned instances having the label $I \in \{1, 2, \ldots, M\}$. Then, we also define a correlation-based metric, which is utilized in our similarity measure since that metric has been previously successfully employed in the literature for analyzing the inter-participants statistical variations (cf section 2). The utilized correlation metric is the summation of two correlation coefficients. The first one \( \rho_{n,1} \) is the correlation between the $n$th participant’s template \( \mathbf{x}_{n}^{w} \) and the new user instance $x$. The second one \( \rho_{n,2} \) is the correlation between the instance $x$ and an artificial reference signal $Y^{x} \in \mathbb{R}^{2N_{h} \times \mathbb{X}}$ that is formed for the $x_{y}$th character, where $N_{h}$ is the number of harmonics (we use 5 harmonics: $N_{h} = 5$). This artificial signal is generated for every character as the composition
of sinusoids of the corresponding tagging frequency harmonics. The definition of this artificial signal can be found in equation 1 of [19]. The first correlation coefficient \( \rho_{n,1} \) can already be regarded as a direct similarity measure between the \( n \)th participant and the new user; however, the second correlation coefficient \( \rho_{n,2} \) (that we want to use) usefully takes into account the correctness of the \( n \)th participant’s DNN’s prediction \( \hat{y}_{n} \).

It is nontrivial to calculate these correlations as the above-defined signals \((x, y, \mathbf{X}^{v})\) are multi-dimensional, hence a reduction is necessary via combinations. Further, though such correlations can model the temporal statistics, the spatial statistics can also be probed using a channel combination \( \mathbf{w}^{(v)} \in \mathbb{R}^{C \times 1} \) selected from the layers of the participants’ DNNs (figure 1). Let the learned channel combination weights from the second layer of the \( n \)th participant’s DNN be denoted by \( \mathbf{w}_{n}(n) \), i.e. \( \mathbf{w}_{n}(n) = \{\mathbf{w}^{(1)}_{n}(n), \mathbf{w}^{(2)}_{n}(n), \ldots, \mathbf{w}^{(N_{h})}_{n}(n)\} \in \mathbb{R}^{C \times N_{h}} \), where \( N_{h} = 120 \) is the number of channel combinations (figure 1). These weights come after the sub-band combination weights \( \mathbf{w}_{I}(n) \), i.e. \( \mathbf{w}_{I}(n) = \{\mathbf{w}^{(1)}_{I}(n), \mathbf{w}^{(2)}_{I}(n), \ldots, \mathbf{w}^{(N_{I})}_{I}(n)\} \in \mathbb{R}^{N_{I} \times 1} \), in the employed DNN architecture. Therefore, we also combine the sub-bands of the new user instance \( x_{v} \) and the \( n \)th participant’s template \( \mathbf{x}_{n}^{v} \) with \( \mathbf{w}_{n}(n) \) in our similarity measure. This outputs \( \mathbf{x}_{n} = \sum_{i=1}^{N_{I}} \mathbf{w}^{(i)}_{I}(n)\mathbf{x}^{(i)}(n) \) for the new user instance \( x \), and similarly \( \mathbf{x}_{n}^{v} \) for the template. Likewise, a harmonic combination \( \mathbf{Y} \in \mathbb{R}^{2N_{I} \times 1} \) is needed in the second correlation coefficient \( \rho_{n,2} \) to combine the harmonics in \( \mathbf{Y}^{v} \). This harmonic combination \( \mathbf{Y} \) is found by CCA (canonical correlation analysis) for the given channel combination \( \mathbf{w}^{(v)} \).

Afterwards, we define the vector of aforementioned correlation coefficients \( \rho_{n}(\mathbf{w}^{(v)}) \) as (‘ is transpose below)

\[
\rho_{n}(\mathbf{w}^{(v)}) = \begin{bmatrix}
\rho_{n,1}(\mathbf{w}^{(v)}) \\
\rho_{n,2}(\mathbf{w}^{(v)})
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\rho((\mathbf{w}^{(v)})^{\top}\mathbf{x}_{n}^{v})(\mathbf{w}^{(v)})^{\top}\mathbf{y}^{v}) \\
\rho((\mathbf{w}^{(v)})^{\top}\mathbf{x}_{n}^{v})(\mathbf{w}^{(v)})^{\top}\mathbf{y}^{v})
\end{bmatrix},
\]

where the rightmost column keeps the regular correlation coefficients between two one-dimensional signals, and the channel combination \( \mathbf{w}^{(v)} \) is obtained from \( \mathbf{w}_{n}(n) \) by maximizing the summation of the squared correlation coefficients:

\[
\mathbf{w}^{(v)} = \arg \max_{\mathbf{w}^{(v)} \in \mathbf{w}_{n}(n)} \sum_{k=1}^{2} (\rho_{n,k}(\mathbf{w}^{(v)}))^{2}. \tag{1}
\]

Based on this setup, we define our similarity measure \( \rho_{n} \) between the \( n \)th participant and the new user as the sum of the squared correlation coefficients as

\[
\rho_{n} = \sum_{k=1}^{2} (\rho_{n,k}(\mathbf{w}^{(v)}))^{2}. \tag{2}
\]

Hence, we obtain the set \( \{f_{\rho(I,j)}\}_{j=1}^{k} \) of most representative \( k \) participants, provided that \( I \) sorts the participants with respect to their similarities to the new user in the decreasing manner, i.e. \( \rho_{I(j_{1})} > \rho_{I(j_{2})} \) if \( j_{1} \geq j_{2} \) and \( I(1) \) is the index of the most correlatively similar participant.

### 4.3. Target identification of the new user instance \( x \)

In our proposed method, the following weighted linear combination of the predictions of the most representative \( k \) participants is the final target identification of the new user SSVEP EEG speller instance \( x \):

\[
y = \arg \max_{i \in \{1, \ldots, M\}} \sum_{j=1}^{k} \rho_{I(j)} \mathbb{I}(y_{I(j)} = i). \tag{3}
\]

The parameter \( k \) (number of most representative participants) can be chosen empirically or by cross-validation using participants’ data. However, since the new user’s statistics are different from the participants’ statistics, using cross-validation might not be reliable. Choosing by hand empirically is not feasible as the best \( k \) could vary for each new user and even for each new user instance. Instead, we propose a more sound strategy. After calculating the predictions for all \( k \in \{1, 2, \ldots, N\} \), we choose the \( k \) of the prediction with the largest confidence that is defined as the weight difference between the most weighted character and the second most. The resulting \( k \) can be different across not only new users but also new users’ SSVEP EEG speller instances. We call this proposed \( k \) selection strategy as **dynamic selection**. We emphasize that (a) the proposed dynamic selection automatically determines the best \( k \) for every new user instance on the fly while identifying the target character, and hence (b) our method removes the difficulty of finding the best \( k \) by not requiring any separate parameter optimization/tuning step (no cross-validation or no manual tuning). This is another important contribution of our study. Finally, (c) by the introduced dynamic selection, we successfully address the person-to-person as well as the within-person statistical variations (from one new user instance to another) in the EEG signals.

We summarize our method **Ensemble-DNN** step by step in algorithm 1. Our **MATLAB code** is available for reproducibility at [https://github.com/osmanberke/Ensemble-of-DNNs](https://github.com/osmanberke/Ensemble-of-DNNs).
Algorithm 1. Our Ensemble-DNN: Target identification of our method for a new user SSVEP EEG speller instance $x$.

1: for $n = 1$ to $N$ do
2:   Get the prediction $\hat{y}_n = f_{(\theta)}(x)$ of the $n$th DNN.
3:   Get the template $\bar{x}_n$.
4:   Get the sub-band combination $\bar{x}_\rho = \sum_{n=1}^{N} w_n^{(\rho)}(x)^{\top}(\rho)$ of the $n$th DNN to $x$.
5:   Get the sub-band combination $\hat{\tau}_\rho = \sum_{n=1}^{N} w_n^{(\rho)}(\tau)^{\top}(\rho)$ of the $n$th DNN to $\hat{x}_\rho$.
6:   Compute the similarity $\rho$ and sort.
7: Compute the $k$ value by using dynamic selection.
8: Compute the final target prediction $\hat{y}$.

5. Performance evaluations

Our performance evaluations are based on the two publically available and widely used benchmark [14] and BETA [15] datasets which were generated with offline SSVEP BCI speller experiments. The benchmark experiments consist of 6 blocks (4 blocks for BETA) and 35 healthy participants (70 participants for BETA). Blocks include spelling of $M = 40$ characters with one trial per each (for both the benchmark and BETA). In the benchmark experiments, the participant was shown a matrix (and BETA). In the benchmark experiments, the participant was shown a matrix (and BETA). In the benchmark experiments, the participant was shown a matrix (and BETA). In the benchmark experiments, the participant was shown a matrix (and BETA).

In each repetition, we only use the data of the new user to check the individual accuracy and ITR. Additionally, we compare our dynamic selection with the regular weighted combination (our Ensemble-DNN without dynamic selection) and the majority voting.

5.1. Results and discussion

On the benchmark dataset (figure 2), our proposed method (Ensemble-DNN) achieves 155.51 bits/min mean ITR (78.61% mean target identification accuracy at 0.9 s of signal length). We demonstrate in figure 2 that our method significantly outperforms all the other compared methods within the range of stimulation of [0.2, 1] s. This is the most used range in the literature, probably because the prolonged exposure to flickering stimulus in the SSVEP experiments quickly becomes too tiring [32]. Hence, the shortest stimulation duration is certainly preferable where one also needs to achieve a sufficiently high ITR. In this sense, our proposed method is highly superior. On the BETA dataset (figure 3), we (Ensemble-DNN) achieve 114.64 bits/min mean ITR (57.95% mean accuracy at 0.7 s). Similar to the benchmark dataset, we demonstrate in figure 3 that our method outperforms the other methods on the BETA dataset as well. To the best of our knowledge, among all the methods in the literature that do not require any sort of user-specific training, data or adaptation from the new user, our proposed Ensemble-DNN are the highest reported performance results to date on these datasets.

Figure 4 shows the mean target identification accuracies for varying $k$ choices (number of selected most representative participants) in our weighted combination in (3). This is our Ensemble-DNN without dynamic selection. Here, the same number $k$ of most representative participants is used for all the new users and for all their SSVEP EEG speller instances. We also compare in figure 4 with the majority voting as an alternative, where the majority of the target character predictions of all the participant-specific fine-tuned DNNs are used (to obtain a final prediction) without weighting. Finally, figure 4 also shows the mean target identification accuracy of our dynamic selection strategy for determining the $k$ value on the fly in the course of making a target character prediction. This is our Ensemble-DNN. Note that in our dynamic selection, the choice for $k$ is automatic and can be different across new users and across their SSVEP EEG speller instances. Hence, our method does also not require any manual parameter tuning or optimization or cross-validation. The results (figure 4) reveal that our proposed dynamic selection strategy performs far better than the weighted combination of no selection and the majority voting, on both the benchmark and BETA datasets.

Table 1 provides comparisons with the no-ensemble approach of global DNN in which we directly use the target character predictions of global DNN without fine tuning (no second stage training) at the participants. In table 1, our proposed ensemble approach (Ensemble-DNN) greatly enhances the no-ensemble approach of global DNN.
in terms of the ITR and accuracy. This superiority is explained in figure 5 which presents two colored dot plots for the benchmark (left) and BETA (right) new users. We clearly observe that lighter colors accumulate on the bottom-right while darker colors accumulate on the upper left parts. This trend indicates the following. If the global DNN accuracy is high (low), then it means most participants are statistically similar (dissimilar) to the new user being tested, and in accordance, our dynamic selection method chooses more (less) participants, with high (low) mean and low (high) standard deviation for $k$, all expected. Consequently, our proposed Ensemble-DNN with dynamic selection relies only on the
Table 1. The mean target identification accuracy (on the first rows) and the mean ITR (on the second rows) comparisons between our proposed Ensemble-DNN and the global DNN (no-ensemble approach) are presented.

| Benchmark | Ensemble | Global | Ensemble | Global |
|-----------|----------|--------|----------|--------|
| 1 sec     | 80.88    | 70.61  | 66.83    | 56.58  |
|           | 151.75   | 123.31 | 113.50   | 88.12  |
| 0.9 sec   | 78.61    | 68.58  | 64.34    | 54.56  |
|           | 155.51   | 126.58 | 114.36   | 89.22  |
| 0.8 sec   | 74.40    | 64.42  | 61.07    | 51.94  |
|           | 153.62   | 123.54 | 114.09   | 89.21  |

Figure 5. Each colored dot represents a benchmark (BETA) new user on the left (right), where the lighter (darker) color indicates higher (lower) global DNN target identification accuracy for that new user. The mean and standard deviation of \( k \) (number of selected most representative participants) are across the SSVEP EEG speller instances of the corresponding new user and based on our proposed dynamic selection.

statistically similar participants and thus outperforms the no-ensemble approach of global DNN.

Another important aspect, along with the performance, is the computational complexity. For a successful implementation in the online setting, the running time of the target identification algorithm must not exceed the break time (algorithm running for each trial must be completed by the end of the break) between two consecutive trials. When our Ensemble-DNN algorithm runs on an Intel Core i7-10 750 CPU, the target identification of one trial takes 0.356 (0.871) s and 0.489 (1.313) s on average for 0.2 s and 1 s of signal lengths, respectively, in the case of the benchmark (BETA) dataset. Since the BETA dataset has a higher number of training participants (69) compared to the benchmark dataset (34), processing BETA trials takes longer time. The most time-consuming part of our target identification algorithm is the channel selection (equation (1)) since the similarity metric is calculated for each training participant as many times as the number of channel combinations (\( N_{\text{ch}} = 120 \)). If the similarity metric can be calculated only once per training participant, then the average running time of our algorithm can be reduced to 0.126 (0.363) s for 1 s of signal length in the case of the benchmark (BETA) dataset. Therefore, to accelerate our algorithm during the similarity metric calculation, one can stop using the unpopular channel combinations (that are statistically less likely to be selected) after processing a few trials from the new user. The same idea can be used for the training participants’ networks as well by dropping the ones that are statistically less likely to contribute to the decisions. Alternatively, by adapting the approach in [33], one can also gradually (across trials) learn a single optimal channel combination for the new user and use only it in the similarity metric calculation for an acceleration. We leave these explorations to a future study. Without any change, our algorithm in its current form is suitable for use in the online setting with around 0.6 s of break time between consecutive trials. We point out that such breaks are already employed, for instance, in the benchmark dataset for allowing the participant some time to blink, relax, reposition, etc.

This study considers the statistical EEG signal variations across persons from the same BCI setting. As there certainly are other sources of EEG variability such as different EEG devices, different tasks or environments, extending our target identification method to wider settings would be valuable which we reserve as another future study.
5.2. Statistical significance analyses
For each $T \in \{0.2, 0.4, 0.6, 0.8, 1\}$, we conduct 5 paired t-tests, pairing our proposed Ensemble-DNN method with the compared methods in figures 2 and 3. We report unadjusted p-values, and the observed difference is called as ‘statistically significant’ (*) if the p-value is less than $\frac{0.05}{5}$ and ‘statistically highly significant’ (***) if the p-value is less than $\frac{0.05}{25}$. For the statistically significant case, we apply single Bonferroni correction by dividing 0.05 by 5, since for each $T$ there are 5 comparisons. And for the statistically highly significant case, we apply double Bonferroni correction by dividing 0.05 by 5, since for each $T$ there are 25 comparisons.

5.2.1. In the case of the benchmark dataset
In terms of the accuracy (figure 2), the least significant difference between our Ensemble-DNN method and the compared methods is observed with (1) FB-C-CNN (** $p = 4.29 \times 10^{-5}$) for $T = 0.2$, (2) FB-C-CNN (** $p = 1.51 \times 10^{-6}$) $T = 0.4$, (3) FB-C-CNN (** $p = 7.83 \times 10^{-6}$) for $T = 0.6$, (4) ttf-CCA ($p = 0.27 \times 10^{-1}$) for $T = 0.8$, (5) ttf-CCA ($p = 0.48$) for $T = 1.0$. For $T = 0.8$, the difference with ttf-CCA is not significant; but it is highly significant (***) with all the others. For $T = 1.0$, the difference with ttf-CCA and FB-C-CCN are not significant; but it is significant (*) with tt-CCA and highly significant (***) with all the others. In terms of ITR (figure 2), the least significant difference between our Ensemble-DNN method and the compared methods is observed with (1) FB-C-CCN (** $p = 4.36 \times 10^{-5}$) for $T = 0.2$, (2) FB-C-CNN (** $p = 9.10 \times 10^{-7}$) for $T = 0.4$, (3) FB-C-CNN (** $p = 2.32 \times 10^{-5}$) for $T = 0.6$, (4) ttf-CCA ($p = 0.18 \times 10^{-1}$) for $T = 0.8$, and (5) ttf-CCA ($p = 0.43$) for $T = 1.0$. For $T = 0.8$, the difference with ttf-CCA is not significant; but it is highly significant (***) with all the others. For $T = 1.0$, the difference with ttf-CCA is not significant; but it is significant (*) with tt-CCA and FB-C-CCN, and highly significant (***) with all the others.

5.2.2. In the case of the BETA dataset
In terms of the accuracy (figure 3), the least significant difference between our Ensemble-DNN method and the compared methods is observed with (1) FB-C-CNN (** $p = 6.68 \times 10^{-8}$) for $T = 0.2$, (2) FB-C-CNN (** $p = 4.42 \times 10^{-5}$) for $T = 0.4$, (3) FB-C-CNN (** $p = 3.18 \times 10^{-5}$) for $T = 0.6$, (4) FB-C-CNN ($p = 0.65$) for $T = 0.8$, (5) FB-C-CNN ($p = 0.55$) for $T = 1.0$. For $T = 0.8$ and $T = 1.0$, the difference with FB-C-CNN is not significant; but it is highly significant (***) with all the others. In terms of ITR (figure 3), the least significant difference between our Ensemble-DNN method and the compared methods is observed with (1) FB-C-CNN (** $p = 6.33 \times 10^{-8}$) for $T = 0.2$, (2) FB-C-CNN (** $p = 8.89 \times 10^{-6}$) for $T = 0.4$, (3) FB-C-CNN (** $p = 3.07 \times 10^{-6}$) for $T = 0.6$, (4) FB-C-CNN ($p = 0.41$) for $T = 0.8$, and (5) FB-C-CNN ($p = 0.35$) for $T = 1.0$. For $T = 0.8$ and $T = 1.0$, the difference with FB-C-CNN is not significant; but it is highly significant (***) with all the others.

6. Conclusion
For SSVEP BCI speller systems to become practical and commonplace, we shall not require any additional data from a new-coming user for algorithm training and system adaptation. However, most existing target identification techniques require that burden of lengthy and tiring EEG data collection processes to achieve high ITR performances. Our proposed method in this work ensures practicality by using the already-existing literature datasets from various previous EEG experiment participants. We first train participant-specific target identifier DNNs based on participants’ own pre-existing data, and then combine the $k$ most representative (in the sense of statistical similarities between the data of participants and the data of new user) DNNs to predict the target character for a given new user instance during a spelling session. Thus, our method does not require any additional training data from a new user who wants to use the system immediately with no hassle. The proposed method achieves significant improvements in the ITR performances compared to the state-of-the-art alternatives while promoting the wide-spread BCI use in daily lives.

Data availability statement
The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/osmanberke/Ensemble-of-DNNs.

Acknowledgment
This work was supported by The Scientific and Technological Research Council of Turkey under Contract 121E452. We thank Can Aksoy, Emirhan Koc, Suayb Arslan and Yigit Catak for their participation in our discussions and their help in our MATLAB implementations and performance evaluations.

ORCID iDs
Osman Berke Guney https://orcid.org/0000-0003-3036-9231
Huseyin Ozkan https://orcid.org/0000-0002-5539-9085

References
[1] Chen Y, Yang C, Ye X, Chen X, Wang Y and Gao X 2021 Implementing a calibration-free SSVEP-based BCI system with 160 targets J. Neural Eng. 18 046094
[2] Xu M, Han J, Wang Y, Jung T-P and Ming D 2020 Implementing over 100 command codes for a high-speed
hybrid brain–computer interface using concurrent p300 and
ssvep features IEEE Trans. Biomed. Eng. 67 3073–82
[3] Cao L, Li G, Xu Y, Zhang H, Shu X and Zhang D 2021 A
brain-actuated robotic arm system using non-invasive
hybrid brain–computer interface and shared control strategy
J. Neural Eng. 18 046045
[4] Zhang R et al 2017 A BCI-based environmental control
system for patients with severe spinal cord injuries IEEE
Trans. Biomed. Eng. 64 1959–71
[5] Mane R, Chouhan T and Guan C 2020 BCI for stroke
rehabilitation: motor and beyond J. Neural Eng. 17 041001
[6] Cheng N, Phua K S, Lai H S, Tam P K, Tang K Y, Cheng K K,
Yew R C H, Ang K K, Guan C and Lim A J H 2020
Brain-computer interface-based soft robotic glove
rehabilitation for stroke IEEE Trans. Biomed. Eng. 67 3339–51
[7] Jin J, Chen Z, Xu R, Miao Y, Wang X and Jung T-P 2020
Developing a novel tactile p300 brain–computer interface
with a cheeks-stim paradigm IEEE Trans. Biomed. Eng. 67
2585–93
[8] Kalagani S P, Laskaris N A, Chatzilari E, Nikolopoulos S and
Kompatsiaris I 2019 A riemannian geometry approach to
reduced and discriminative covariance estimation in brain
computer interfaces IEEE Trans. Biomed. Eng. 67 246–55
[9] Chen X, Wang Y, Gao S, Jung T-P and Gao X 2015 Filter
bank canonical correlation analysis for implementing a
high-speed SSVEP-based brain-computer interface J. Neural
Eng. 12 046008
[10] Yin E, Zhou Z, Jiang J, Yu Y and Hu D 2015 A dynamically
optimized ssvep brain–computer interface (BCI) speller
IEEE Trans. Biomed. Eng. 62 1447–56
[11] Guney O B, Oblokalov M and Ozkan H 2022 A deep neural
network for ssvep-based brain-computer interfaces IEEE
Trans. Biomed. Eng. 69 932–44
[12] Zerafa R, Camilleri T, Falzon O and Camilleri K 2018 To
train or not to train? a survey on training of feature
extraction methods for SSVEP-based BCIs J. Neural Eng. 15
051001
[13] Nakain ohi M, Wang Y, Chen X, Wang Y, Gao X and Jung T
2018 Enhancing detection of ssveps for a high-speed brain
speller using task-related component analysis IEEE
Trans. Biomed. Eng. 65 104–12
[14] Wang Y, Chen X, Gao X and Gao S 2017 A benchmark
dataset for ssvep-based brain–computer interfaces IEEE
Trans. Neural Syst. Rehabil. Eng. 25 1746–52
[15] Liu B, Huang X, Wang Y, Chen X and Gao X 2020 Beta: a
large benchmark database toward SSVEP-BCI application
Front. Neurosci. 14 627
[16] Guney O B, Koc E, Aksoy C, Catak Y, Arslan S and
Ozkan H 2021 Adaptive boosting of dnn ensembles for
brain-computer interface spellers Signal Processing and
Communications Applications Conf. (SUI) pp 1–4
[17] Yijun W, Ruiping W, Xiaorong G and Shangkai G 2005
Brain-computer interface based on the high-frequency
steady-state visual evoked potential Int. Conf. on Neural
Interface and Control pp 37–39
[18] Friman O, Volosyak I and Graser A 2007 Multiple channel
detection of steady-state visual evoked potentials for
brain–computer interfaces IEEE Trans. Biomed. Eng. 54
742–50
[19] Lin Z, Zhang C, Wu W and Gao X 2006 Frequency
recognition based on canonical correlation analysis for
ssvep-based BCIs IEEE Trans. Biomed. Eng. 53 2610–4
[20] Nan W, Wong C M, Wang B, Wan F, Mak P U, Mak P L and
Vai M I 2011 A comparison of minimum energy
combination and canonical correlation analysis for SSVEP
detection IEEE/EMBS Int. Conf. on Neural Engineering pp
469–72
[21] Yuan P, Chen X, Wang Y, Gao X and Gao S 2015 Enhancing
performances of SSVEP-based brain–computer interfaces via
exploiting inter-subject information J. Neural Eng. 12
046006
[22] Waytowich N R, Faller J, Garcia J O, Vetel J M and Sajda P
2016 Unsupervised adaptive transfer learning for steady-state
visual evoked potential brain–computer interfaces IEEE Int.
Conf. on Systems, Man and Cybernetics (SMC) pp 483–90
[23] Waytowich N, Lawhern V J, Garcia J O, Cummings J, Faller J,
Sajda P and Vettel J M 2018 Compact convolutional neural
networks for classification of asynchronous steady-state
visual evoked potentials J. Neural Eng. 15 066031
[24] Ravi A, Beni N H, Manuel J and Jiang N 2020 Comparing
user-dependent and user-independent training of cnn for
SSVEP BCI J. Neural Eng. 17 026028
[25] Zhao D, Wang T, Tian Y and Jiang X 2021 Filter bank
convolutional neural network for SSVEP classification IEEE
Access 9 147129–41
[26] Lawhern V J, Solon A J, Waytowich N R, Gordon S M,
Hung C P and Lance B J 2018 Egnet: a compact
convolutional neural network for eeg-based brain–computer
interfaces J. Neural Eng. 15 056013
[27] Rodrigues P L C, Junten C and Congedo M 2019 Riemannian
procrustes analysis: transfer learning for brain–computer
interfaces IEEE Trans. Biomed. Eng. 66 2390–401
[28] Wu D, Xu Y and Lu B-L 2022 Transfer learning for eeg-based
brain–computer interfaces: a review of progress made since
2016 IEEE Trans. Cogn. Develop. Syst. 4 4–19
[29] Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G
and Vaughan T M 2002 Brain–computer interfaces for
communication and control Clin. Neurophysiol. 113 767–91
[30] Zheng X, Xu G, Zhang Y, Liang R, Zhang K, Du Y, Xie J and
Zhang S 2020 Anti-fatigue performance in ssvep-based
visual acuity assessment: a comparison of six stimulus
paradigms Front. Hum. Neurosci. 14 301
[31] Wong C M, Wang Z, Nakain ohi M, Wang B, Rosa A, Chen P,
Jung T-P and Wan F 2022 Online adaptation boosts
ssvep-based bci performance IEEE Trans. Biomed. Eng. 69
2018–28