IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received September 9, 2020, accepted September 17, 2020, date of publication September 28, 2020, date of current version October 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3027429

Electroencephalogram-Based Preference Prediction Using Deep Transfer Learning

MASHAEL S. ALDAYEL[1,3], **(Member, IEEE), MOURAD YKHLEF**[1], **AND ABEER N. AL-NAFJAN**[2]

[1]Information System Department, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[2]Computer Science Department, College of Computer and Information Sciences, Al-Imam Mohammad Ibn Saud Islamic University, Riyadh 11432, Saudi Arabia
[3]Information Technology Department, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

Corresponding author: Abeer N. Al-Nafjan (annafjan@imamu.edu.sa)

This work was supported by the Deanship of Scientific Research through the Initiative of DSR Graduate Students Research (GSR) Support, King Saud University.

ABSTRACT Transfer learning is an approach in machine learning where a model that was built and trained on one task is re-purposed on a second task. The success of transfer learning in computer vision has motivated its use in neuroscience. Although common in image recognition, the use of transfer learning in EEG classification remains unexplored. Most EEG-based neuroscience studies depend on using traditional machine learning algorithms to answer a question, rather than on improving the algorithms. Developing algorithms for transfer learning for EEG can also assist with problems of low data availability in EEG classification. The primary objective of this study is to investigate EEG-based transfer learning and propose deep transfer learning models to transfer knowledge from emotion recognition to preference recognition to enhance the classification prediction accuracy. To the best of our knowledge, this is the first study demonstrating the effect of applying deep transfer learning between EEG-based emotion recognition and EEG-based preference detection. We propose different approaches for deep transfer learning models to detect preferences from EEG signals using the preprocessed DEAP dataset. Two types of features were extracted from EEG signals, namely the power spectral density and valence. We built three models of deep neural networks: basic without transfer learning, fine-tuning of deep transfer learning, and retraining of deep transfer learning. We compared the performance of deep transfer learning with those of deep neural networks and other conventional classification algorithms such as support vector machine, random forest, and k-nearest neighbor. Although the deep neural network classifiers achieved a high accuracy of greater than 87%, deep transfer learning achieved the highest accuracy result of 93%. The results demonstrate that although the proposed deep transfer learning approaches exhibit higher accuracy than the support vector machine and k-nearest neighbor classifiers, random forest achieves results similar to those of deep transfer learning.

INDEX TERMS Data mining, brain-computer interfaces, emotion recognition, supervised learning, artificial neural networks, signal processing, consumer behavior.

I. INTRODUCTION

Human variability induces a variability in the related EEG across subjects or sessions and even time within a subject. Non-stationary EEG signals create a need for calibration to overcome inconsistency in the distinctive classification label problems. Transfer learning (TL) can solve a task by utilizing knowledge acquired when learning another different but related task. This includes methods designed to enhance the performance of a classifier trained on a particular task, session, or subject, depending on the information acquired while learning a related task [1]–[3].

TL can relax the EEG-based brain-computer interface (BCI) technology limitations by obviating the need to calibrate from the starting point, decreasing noisy transferred data, and relying on the previously available data to increment the EEG data sizes [1].

TL techniques involve the use of one dataset to create an initialization for the classification of another dataset. They can use a model trained on one dataset as an initialization

The associate editor coordinating the review of this manuscript and approving it for publication was Yiming Tang.

for a model that is trained on another dataset. TL can be implemented in BCIs to transfer information between tasks, subjects, or sessions. Subject-to-subject transfer of the same tasks is most commonly investigated in applying TL to BCIs and particularly to EEG-based BCIs [1], [2].

In recent TL research [4], task-to-task approaches have achieved improved results using labeled data from the source domain to study classifiers for the destination domain. Most research on TL in BCIs focused on the transfer of the information across subjects or sessions, but task TL remains mostly unexplored [3]. According to a recent review [3], there have been no studies reported on task TL in affective BCIs that identify emotion or preferences from EEG. In this work, we transfer knowledge of the task identification between emotion and preferences. We apply task-to-task TL on a single-trial basis to find common discriminative information of preferences utilizing domain adaptation.

Furthermore, EEG-based studies for detecting consumer preferences are in a very early phase, although several studies have been conducted on EEG-based emotion recognition [5]. According to Teo *et al.* in [6], preference classification is more challenging than other types of emotion classification due to its comparatively weaker induction capacity. For example, strong emotions such as anger are induced more powerfully [7].

In this study, our primary objective is to investigate EEG-based TL and propose deep transfer learning (DTL) model to transfer knowledge from emotion recognition to preference recognition with the aim of enhancing the accuracy of classification prediction. We mainly investigated the relationship between EEG classification of preference and emotion at a very deep level. we compared the performance of deep learning with deep neural networks (DNNs) and other conventional classification algorithms such as support vector machine (SVM), random forest (RF), and k-nearest neighbor (KNN). The main research questions of this study were as follows: How can knowledge be transferred between the emotion domain and the preference domain? Can TL improve the performance of deep learning?

The rest of this paper is organized as follows: Section II reviews some of the background in the field of preference recognition. Sections III and IV introduce the principal concepts of TL and DTL, respectively, in detail. Section V describes our proposed DTL, the EEG dataset and experimental details. Section VI discusses the results, and Section VII states the conclusions. Finally, Section IX suggests the possible future directions for research.

II. PREFERENCE RECOGNITION BACKGROUND

This section provides a detail description of EEG-based preference recognition, specifically the neural correlation of preference measurements with the cognitive and affective perspectives. Preference is an evaluative judgment in term of liking or disliking a set of objects. [8]. Although several studies [6], [9]–[11] have found that EEG can recognize preferences, understanding the psychological process underlying the measurement of preferences is important [12].

EEG-based preferences can be measured and defined from either cognitive or affective perspectives, as shown in Figure 1. These perspectives reflect discrimination between the wanting and liking processes in neuroscience research. Wanting occurs unconsciously and is measured in terms of changes in behavior, arousal, or eye fixation. It is reflected in brain activation in the basal ganglia and the nucleus accumbens. By contrast, liking is related to the overt and conscious hedonic experience, measured through explicit preference statements, and is reflected in the brain activation of prefrontal regions, such as the orbitofrontal cortex and potentially the anterior insula [8]. The following subsection describes the neural correlations of the EEG-based preferences from the affective and cognitive perspectives.

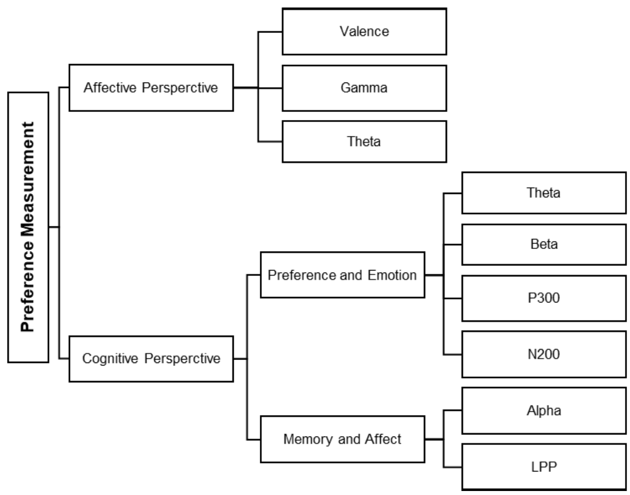

FIGURE 1. Preference measurements from affective and cognitive perspectives.

A. COGNITIVE PROCESS AND PREFERENCES

Choice raises the idea of a fixed selection of a desired service or product, which is driven by interior likes and dislikes, or other preferences. Although such interior processes basically overlap with the field of cognitive psychology, understanding the cognitive processes before and after decision-making has the potential to further develop the field of marketing research [13]. Several EEG studies have emphasized evaluating cognitive and emotional perception in reaction to a certain event or stimulus.

Based on the neuromarketing studies reviewed in the literature, researchers have been using EEG and ERP to understand how one or more of the cognitive activities (attention, memory, preferences, emotions) relate to ads, brands, products, or prices [9]. The relationship between cognitive processes and consumer preferences is described in detail in this section.

N200 and P300 components and theta and beta waves have been used in marketing studies for cognitive processes

TABLE 1. Neuromarketing studies in cognitive processes associated with emotions and preferences.

Component	Related Research
N200	Regarding extended branding and product class [14], the N200 component was unconsciously enhanced when facing a new outfit brand, particularly under the effect of informed negative emotions. The Cerebro system [15] combined the usage of the N200 mean, N200 minima, and ERSP to rank products according to customer preferences. Another study [16] has similarly discovered that a minor N200 exposes value-related cognitive techniques. These studies show implied preferences for products using N200.
P300	P300 can be produced with conscious tasks to assess brand extension since it is associated with the decision-making attention given to products [17]. In addition, P300 components can predict preferences for product features, as observed when an improved P300 was detected after consumers saw a preferred product [18].
Beta	Greater beta waves (16–18 Hz), which were linked with preference and reward processes, were distinguished when consumers watched movie trailers [11], [19], and a higher amplitude of the beta band (16–18 Hz) corresponded to higher consumer ranking of the movie. Moreover, they found an association between the population preferences for the movies and frequency oscillations in the gamma band (60–100 Hz).
Theta	Greater theta bands are prompted when watching TV commercials that were ranked as pleasant (like) versus unpleasant (dislike) [20]. Telpaz [16] has similarly discovered that a lower theta frequency is associated with failures and negative results toward products. In addition, a lower theta frequency has been linked to effective anti-smoking public service announcements [21] and viewing foreign products compared to local products [22].
Alpha	The frontal alpha brainwaves show the presence of pleasant emotional responses to TV ads [19], [23]–[25] as displayed in the hemispheric asymmetry of the brain. The frontal alpha asymmetry should be qualified in the context of emotion recognition in response to ad stimuli [26]. Touchette et al. [27] found frontal asymmetry in the alpha band linked to consumers' unconscious reaction to product attractiveness at electrodes F3 and F4. Similarly, Modica et al. [22] linked higher alpha to comfort food and foreign food products. An award-winning campaign (receiving prizes) in anti-smoking public service announcements has been linked to a higher alpha band [21].
LPP	The LPP component, correlated with emotion, was enhanced when shopping for luxury brands compared to non-luxury brands. LPP was particularly noticeable when decisions were made in the context of a group setting, indicating that social pressure influences LPP [9] [26].

associated with emotion and preferences. Alpha waves and the late positive potential (LPP) have been used to associate cognitive processes with affect and memory. The studies of each of these are listed in Table 1.

B. AFFECTIVE PERSPECTIVE OF PREFERENCES

This section explains preferences from the affective and emotional perspectives. First, we explain emotion modeling and classification. Then, we illustrate how emotion classification has been used for preference classification.

1) EMOTION MODELING AND CLASSIFICATION

Emotion models are typically categorized as of two types, namely, discrete and dimensional. Discrete emotion models classify emotions into a limited number of separate states based on physiology and neural expressions. Most studies have recognized six fundamental emotions of disgust, happiness, surprise, anger, fear, and sadness. Dimensional emotion models discriminate emotional states using different dimensions, mostly valence and arousal. Valence is the degree of pleasantness linked with an emotion, and arousal refers to the strength of an emotion. Most research has used dimensional models for emotion classification [7]. Some emotion recognition research [28], [29] has used a binary classification of emotion into positive and negative based on dimensions such as valence, arousal, and domination.

Emotion can be measured using many instruments in terms of features such as blood pressure, skin conductance, heart rates, and brain waves, the latter observed using EEG. Classification of EEG-measured emotion typically includes transforming EEG signals into features fed to data mining algorithms trained on labeled data to anticipate emotion. Research has shown that emotion classification is reliably produced using EEG [6], [7].

2) EMOTION CLASSIFICATION OF PREFERENCES

Preference classification can be defined as a subdomain of emotion classification. It specifically detects a user's like or dislike when affected by a stimulus. Moreover, preference classification is more challenging than other types of emotion classification due to its comparatively slight induction, e.g., emotions such as anger are more powerfully induced [6].

Preferences are related to positive emotion. Since neuropsychological studies have confirmed the relationship between EEG data and emotions [5], [7], [30], this relationship can be utilized in preference detection. Emotions are typically categorized using a bi-dimensional valence-arousal approach. Such a model represents emotional states using two

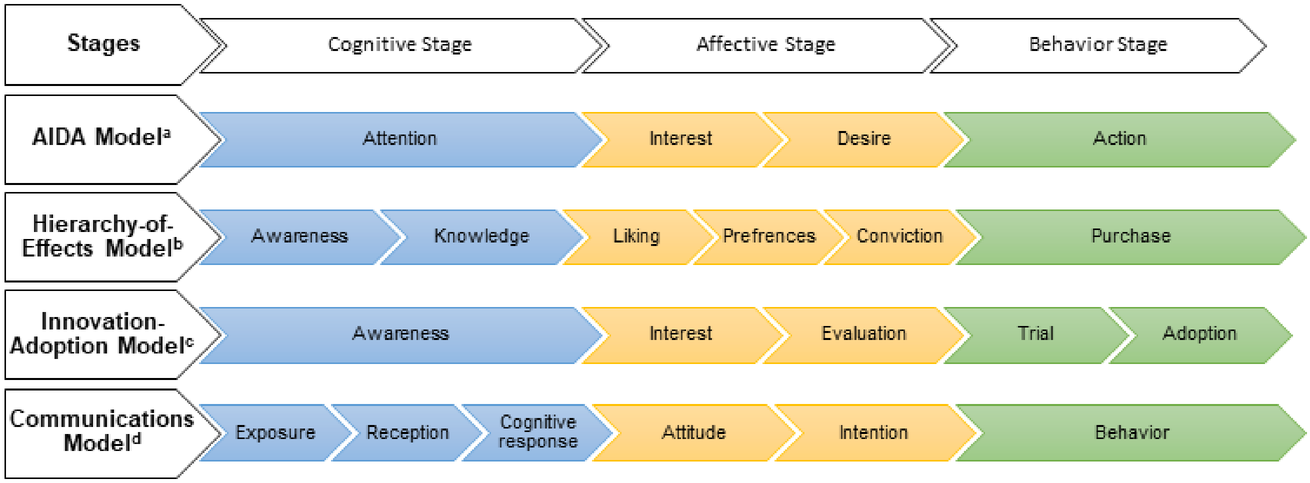

FIGURE 2. Response Hierarchy Models [33] : (a) E. Strong, The Psychology of Selling (New York: McGraw-Hill, 1925), p. 9; (b) R. Lavidge and G. Steiner, "A Model for Predictive Measurements of Advertising Effectiveness," Journal of Marketing (October 1961), p. 61; (c) E. Rogers, Diffusion of Innovation (New York: Free Press, 1962), pp. 79–86; (d) D. Vakratsas and T. Ambler, "How Advertising Works: What Do We Really Know?" Journal of Marketing (January 1999), pp. 26-43.

dimensions, arousal and valence; arousal is the strength of an emotion (high or low concentration), and valence is the direction of an emotion, positive or negative.

Music preference studies [31] have linked higher preference for products with positive valence and high/low arousal. The degree of arousal depends on other factors such as age, gender, and uncertainty. Ramsoy *et al.* [8] found that higher arousal and lower preference ratings are linked with uncertain perception of brand logos. They verified this relationship using various stimuli (music and pictorial art). Michael *et al.* [32] investigated the emotional reaction (arousal and valence) of tourism preferences using various stimuli (word, image, and video). They found that images had higher affective responses than words in travel decision-making driven by unconscious preference. Most approaches used for EEG-based emotion recognition depend on assessments of time and frequency [5].

In time-based assessment, event-related potential (ERP) components reveal emotions based on the representation of the bi-dimensional valence-arousal approach. The ERP components of short to middle expectancies are associated with valence, whereas the ERP components of middle to long expectancies are associated with arousal. In frequency-based assessment, the power of the frequency ranges has been linked to diverse emotions (happy, sad, angry, fearful, or neutral). The stimulus can modify the spectral power synchronization of frequency ranges [5]. Gamma bands are associated with a happy emotional state, while the theta band is associated with changes in the emotional state. To summarize, both approaches (time- and frequency-based) can be used to recognize customer preferences and emotional states in neuromarketing.

Figure 2 shows the most popular consumer response models. All of these models link the affective stage with liking and preferences, justifying the use of affective pictures and datasets in neuromarketing research. Even though several studies have evaluated EEG-based emotion recognition, the EEG-based studies for preference detection in customers are in a very early phase. In addition, preference classification is considerably more challenging than other types of emotion classification due to its relatively weak re-creation [6].

III. TRANSFER LEARNING (TL)

TL can be defined as a regularizer for solving a particular task by passing knowledge from the origin domain to the destination domain, as shown in Figure 3. This section describes the TL categorizations, approaches, and transferred information categories in detail.

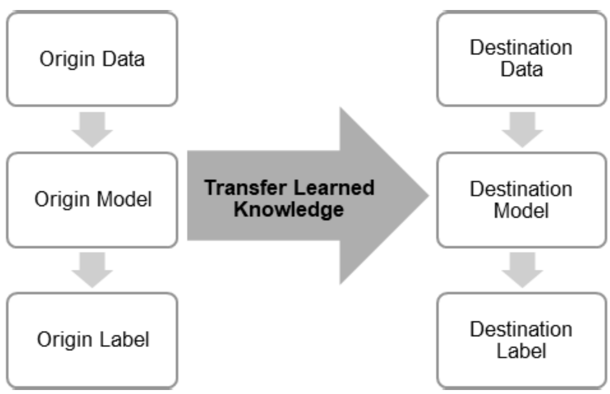

FIGURE 3. Concept of transfer learning.

A. TL CATEGORIZATIONS

In TL, the origin domain is always known, while the destination domain can be known (inductive) or unknown (transductive) [1]. Accordingly, TL can be classified based on the relationship between the origin and destination domains into three subcategories of inductive, transductive, and unsupervised.

1) **Inductive TL**
In inductive TL, the destination task is discovered according to the knowledge transmitted from the source task. The tasks can vary between the origin and destination regardless of their domains. The availability of class labels in the destination domain is required in inductive TL to enhance the prediction estimation in the destination domain. Inductive TL can be subclassified based on the availability of class labels in the origin domain. If accessible, multitask learning for the origin and destination tasks is used at the same time. Alternatively, self-taught learning is used [1], [34].
2) **Transductive TL**
The origin task is the same as the destination task regardless of their domains in transductive TL. The availability of class labels in the origin domain is required to enhance the prediction estimation in the origin domain. Transductive TL can be subclassified based on the degree of similarity of the feature spaces between the origin and destination domains. If they are identical, then a homogeneous TL method, such as sample selection bias, is used. Alternatively, a heterogeneous TL method is used, such as domain adaptation [1], [34].
3) **Unsupervised TL**
In unsupervised TL, both the origin and destination tasks are relevant but different, and the class labels are missing in both the origin and destination domains. The purpose of unsupervised TL is to determine the clustering, estimate the density, and reduce dimensional tasks in the destination domain. [1], [34].

B. TRANSFERRED INFORMATION CATEGORIES
Two categories of information can be transferred in BCIs, discriminative or stationary information. The category suitable for the transferred information is determined by the similarity between the origin and destination domains. If the two domains are similar and the dataset is limited, then transferring discriminative information is more appropriate for highlighting invariable features and models. However, if the domains vary sufficiently, then transferring stationary information is more appropriate for constructing more invariant features or models based on the common information across different domains [1], [34].

C. TL APPROACHES
TL strategies and approaches can be classified based on the transferred information category into the three subclasses of feature-representation, instance-based, and classifier-based TL. [34], [35].

1) **Feature-representation TL**
This approach allows the reconstruction of the features for the destination domain using knowledge from the origin domain. The information transferred across domains is reconstructed as a new feature representation allowing the generalization of the classifiers on the destination test data and reducing the error rate. In BCI research, spatial filters, such as common spatial patterns (CSPs), are used to extract the features from EEG data.
2) **Instance-based TL**
This approach allows the reuse of instances from the origin domain to assist in the learning of the destination domain. The well-known techniques using this approach for BCIs are instance reweighting, which assigns weights to instances from the origin domain for usability, and importance sampling, in which some values have a stronger effect on the learning process.
3) **Classifier-based TL**
This approach allows the reuse of a classifier from the origin domain to assist in the learning of the destination task. It can take two forms, namely, ensemble learning and domain adaption of classifiers. Ensemble learning of the classifiers merges many classifiers across many domains to enhance the prediction accuracy of the destination domain. Domain adaptation allows the reuse of classifiers from the origin domain through the adjustment of the parameters related to the destination domain. It can deal with data changes between the origin domain and the destination domain, but both domains must share some common information. This method is commonly used for transferring discriminative and stationary information across sessions [34], [35].

IV. DEEP TRANSFER LEARNING (DTL)
Implementation of TL is aimed at enhancing the knowledge generalization capacity for machine learning, and implementation of deep learning is aimed at facilitating rapid reprocessing of data through re-engineering of features and extraction of high-level abstract features with undetectable dependencies at the same time. In BCIs, deep learning operates directly on the raw EEG signals to learn unique information through backpropagation in neural network structures. Generally, the goal of implementing DTL as a classifier is to reduce the training time and enhance the accuracy compared with isolated DNN and TL [1]

DTL has recently been categorized into four types of mapping-based, instance-based, adversarial-based, and network-based, as show in Figure 4 [1], [36].

1) **Instance-based**
Instance-based DTL selects particular instances from the origin domain by adjusting the weights. These instances are assigned suitable weight values and

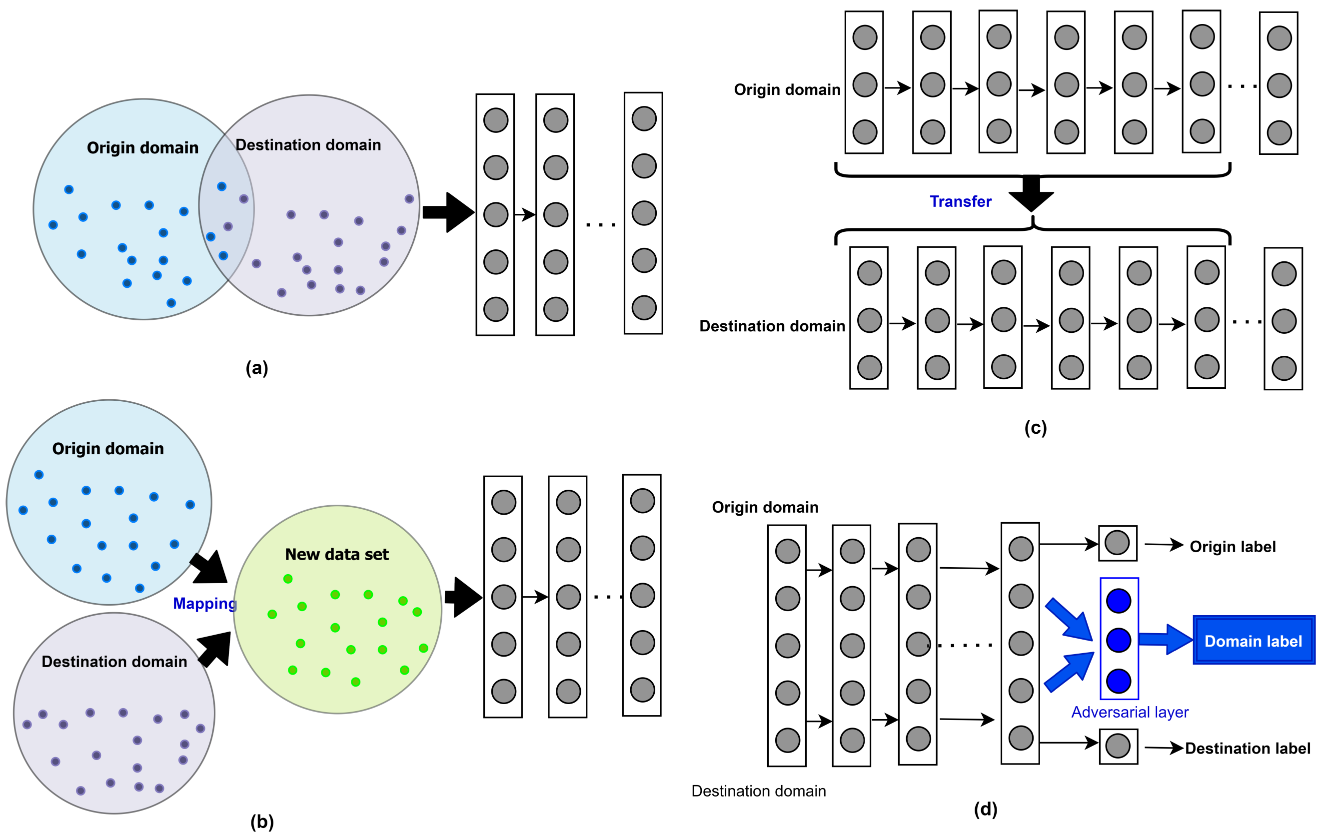

FIGURE 4. Four categories of deep transfer learning (DTL): (a) instance-based, (b) mapping-based, (c) network-based and (d) adversarial-based DTL.

are used as the training set in the destination domain [1], [36].

2) **Mapping-based DTL**
Mapping-based DTL combines instances by matching similarities between the origin and destination domains. The combined dataset includes instances from both domains that are similar and appropriate for a unified DNN [1], [36].

3) **Adversarial-based DTL**
Adversarial-based DTL uses adversarial techniques such as generative adversarial networks (GANs) to deceive models using malicious input. The goal of adversarial techniques is to find transferable feature representations that are appropriate for both the origin and destination domains in distinguishing the learning task [1], [36].

4) **Network-based DTL**
Network-based DTL allows the reuse of the elements of the network structure and parameters. Such a network is pretrained in the origin domain for reuse in the destination domain. [1], [36].

V. LITERATURE REVIEW

We classified the main techniques of DTL in EEG into the image and model-based approaches. The first approach is based on converting the EEG signals to spectrograms (i.e., time-frequency spectrum images). Such an implementation [37]–[39] benefits from research on image recognition through the direct use of the same well-defined pretrained convolutional neural network (CNN) models such as VGG16, AlexNet, and Inception-v3. Xu *et al.* in [37] proposed a deep transfer neural network classifier for EEG motor imagery using VGG-16. Raghu *et al.* [38] proposed similar TL approaches using ten well-known pretrained CNNs, such as AlexNet, VGG16, and GoogLeNet, to identify optimal networks in EEG seizure classification. Similarly, Tan *et al.* in [39] proposed a deep transfer neural network classifier for EEG music imagery using VGG-16, VGG-19, ResNet and AlexNet.

Since EEG signals typically show high intra- and inter-individual variability as well as a low signal-to-noise ratio [40], the second TL approach makes use of the transferability of knowledge between different trained models over different EEG experiments. This can be achieved by adapting pretrained models of deep learning between subjects (i.e., intra-experimental transfer) and between experiments (i.e., inter-experimental transfer). Uran *et al.* in [40] used a TL approach between experiments and demonstrated the ability to improve the performance of classification models trained with limited quantities of EEG data, leading to highly accurate results and faster convergence when training another model. They also found that the best accuracy was achieved when a model is retrained with frozen lower layers (i.e., the same hyperparameter values). Another TL approach was used [2] between subjects through two-step training: (1) training the model across different subjects to have a unified model for all but one subject and then (2) customizing the model based on a specific subject through weight initialization of the unified model. Such a fine-tuning approach to TL has improved the accuracy of all the models. It also reduced overfitting, which is a common problem in DNNs. William [2] also found that deep learning with long short-term memory units (LSTMs) outperforms other conventional methods. In recent TL research [4], task-to-task approaches have achieved improved results using labeled data from the source domain to study classifiers for the destination domain. Most TL research in BCIs focused on transferring information across subjects or sessions, but task TL remains mostly unexplored [3]. According to a recent review [3], no studies have been reported on task TL in affective BCIs that identify emotion or preferences from EEG. In this research, we transfer the knowledge of the task identification between emotion and preferences. We apply task-to-task TL on a single-trial basis to find common discriminative information of the preferences utilizing domain adaptation.

VI. PROPOSED DEEP TRANSFER LEARNING

In this section, we describe the methodology that we used in implementing DTL for EEG-based preference detection. We begin by describing the DEAP benchmark dataset and feature extraction. Then, we illustrate the classification models and propose different DNN models between the emotion and preference domains.

A. DATASET DESCRIPTION

DEAP [41] is a benchmark database developed for affective analysis. Several studies [42]–[44] have adopted the DEAP dataset in preference classification, which is a subdomain of emotion classification. We excluded the self-assessment "like" rating in the DEAP dataset, as the data owners [41] reported conflicting findings between the activation in the left alpha power, reflected in a high valence with a subjective "like" rating that was not consistent with that of similar studies. Therefore, we considered the valence as a preference indicator to identify the true preferences, i.e., like or dislike. We used the self-assessment valence in a 9-point Likert scale and converted it as follows: (1) dislike if the rating of the valence ranged between zero and five and (2) like if the rating of the valence was above five. We preprocessed the EEG dataset, down-sampled the date to 128 Hz, applied a (4.0-45.0 Hz) bandpass filter, and eliminated EOG artifacts with ICA. Finally, we selected the channels of AF3, AF4, F3, F4 and Fz to reduce the computational overhead of the subsequent steps.

B. FEATURE EXTRACTION

The power spectral density (PSD) was used to extract the bands from the EEG signals. Then, the extracted bands were used to calculate the valence, which was chosen to measure the preference in this study because our previous work in [45] showed that valence contains sufficient information to achieve the highest classification accuracy. We applied various valence equations and investigated the relationship with the DEAP self-assessment valence measurement. The computation of the valence is described in our previous report [45].

C. PREFERENCE CLASSIFICATION

Our study aimed to classify the EEG signals into two preference states, like or dislike. We proposed DTL classifiers and compared their performance with those of the DNN, SVM, KNN and RF classifiers. The block diagram of the proposed DNN classifier is provided in our previous work [45]. However, we changed the hyperparameters in each DNN classifier. We conducted a two-fold experiment: one with a DNN classifier built from scratch and the other using a DTL classifier adapted from emotion recognition.

1) DEEP NEURAL NETWORK (DNN) CLASSIFIER

In our previous work [45], we used a DNN architecture in our experiments that was a fully connected feed-forward neural network with three hidden layers containing units involving rectified linear activation functions (ReLu). The output is obtained as a softmax layer with a cross-entropy cost function or a tan layer with a hinge cost function. The input layer contains 2,367 units, and 75% of the units in each hidden layer are from its predecessor (previous) layer. In particular, the first, second, and third hidden layers involve 1,800, 1,300, and 800 units, respectively. The output layer dimensions pertain to the number of target preference state (2) units. To train the DNN classifier, we used Adam gradient descent with three objective loss functions, namely, the binary cross-entropy, categorical cross-entropy, and hinge cross functions. For hyperparameter tuning, we considered reasonable defaults and followed established best practices, with a start learning rate of 0.001. Then, we reduced the rate linearly with each epoch such that the learning rate for the last epoch was 0.0001. We set the dropout for the input and hidden layers as 0.1 and 0.05, respectively. The stopping criterion of the network training was determined according to the model performance on a test set. If the network began to overfit, then the network training was stopped. This stopping criterion was helpful for reducing the possibility of overfitting of the validation data. The network was tested on a test set that contained approximately 20% of the data samples in the dataset.

2) DTL CLASSIFIER

We use DTL of emotion-recognition [30] using two approaches, namely, retraining and fine-tuning of hyperparameters.

1) **Retraining approach of DTL**
   In the first approach, we use pretrained weights as the starting point. Considering the similarity between

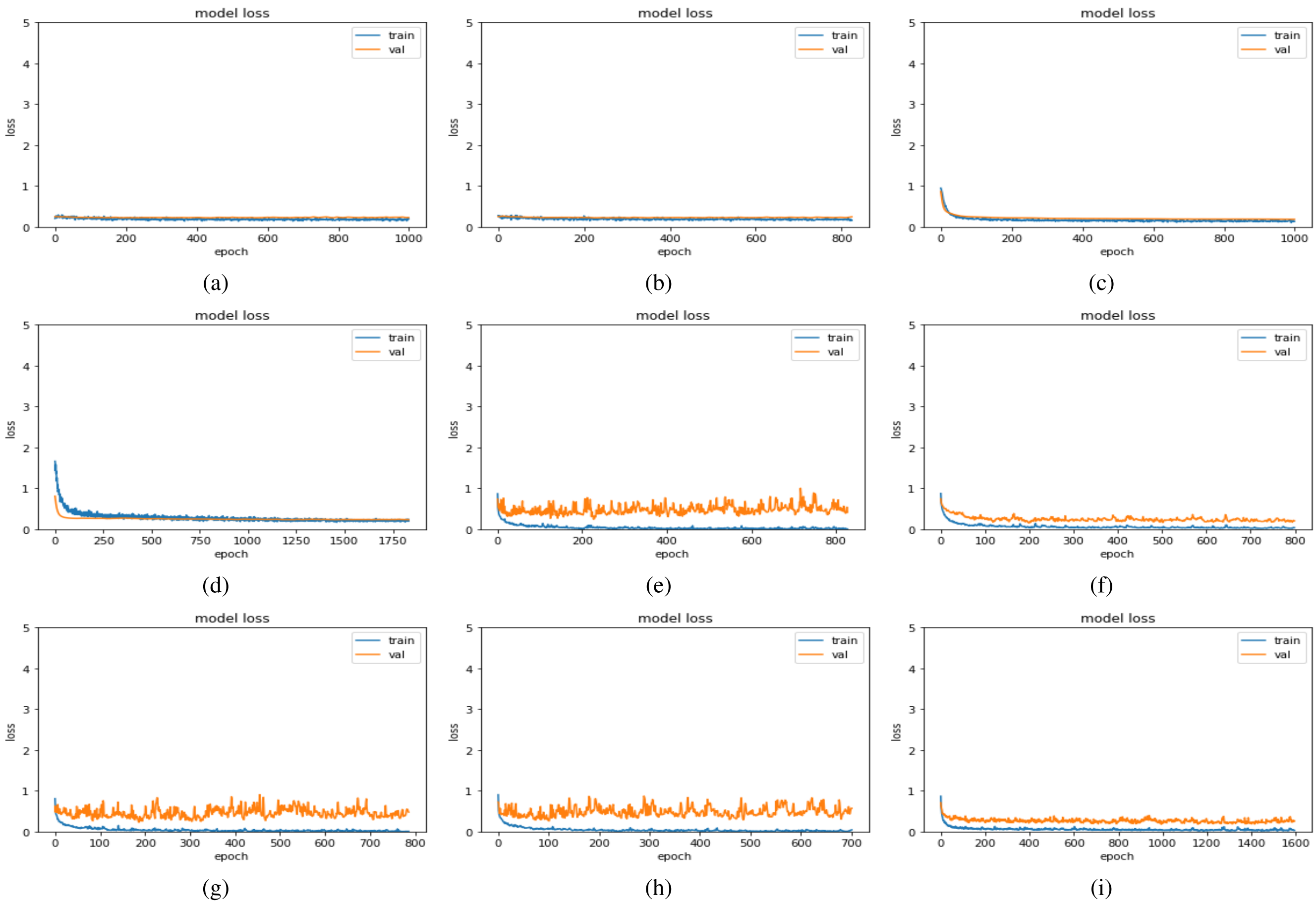

FIGURE 5. **Loss per epoch on the training and validation sets in DTL with the retraining approach of TL (top three charts), fine-tuning approach of TL (middle three charts), and basic DNN without TL (bottom three charts) using different cross-entropy functions: (a,d,c) categorical cross-entropy, (b,e,h) binary cross-entropy, and (c,f,i) hinge function.**

preferences and emotion in using valence as the measurement, we remove the last fully connected layer and replace it with a layer matching the number of classes, like or dislike, in the DEAP dataset. We initialize the weights randomly in the new fully connected layer but initialize the remainder of the weights using the pretrained weights of an emotion-recognition DNN. Finally, we retrain the entire neural network. Because the original training set and the new dataset share higher level features, the entire neural network is also used.

2) **Fine-tuning approach of DTL**
In the second approach, we slice off the end of the emotion DNN and add a new fully connected layer that matches the number of classes, like or dislike, in the DEAP dataset. Then, we randomize the weights of the new fully connected layer but freeze all the weights from the pretrained emotion DNN. Finally, we train the network to update the weights of the new fully connected layer. To avoid overfitting, the weights of the original DNN are held constant without retraining.

VII. RESULTS AND DISCUSSION

This study investigated the application of machine learning and computational statistics in consumer preference (like or dislike) prediction using the different DNN, RF, KNN, and SVM classification algorithms. We used different evaluation measurements of accuracy, recall, and precision. The accuracy was calculated as the average of the binary measurements in which the score of every class was weighted by its availability in the real data. Precision is the proportion of (like) preference predictions that were actually correct. Recall is the proportion of actual (like) preferences that were predicted successfully.

A. DTL AND DNN

To evaluate the performance of the DTL and DNN classifiers, we used holdout (train/test splitting) as a basic validation approach. We compared the performance of the DTL and DNN classifiers with and without TL, respectively. We used two TL approaches, namely, retraining and fine-tuning of the hyperparameters. The accuracy, recall, and precision results of the proposed DNN classifiers are presented in Table 2 with

TABLE 2. Results of preference recognition of DNN classifiers with and without (w/o) TL adapted from emotion recognition using holdout validation and different cross-validation functions.

	Hinge cross			Binary cross			Categorical cross		
Transfer from Emotion	w/o TL	with TL		w/o TL	with TL		w/o TL	with TL	
		Fine-tuning	Retraining		Fine-tuning	Retraining		Fine-tuning	Retraining
Accuracy	90%	93%	92%	87%	93%	92%	90%	91%	93%
Recall	90%	93%	92%	87%	93%	92%	90%	91%	93%
Precision	90%	93%	92%	89%	94%	92%	91%	92%	93%
Loss rate	56%	20%	18%	57%	24%	24%	47%	28%	22%
Batch size	256	256	1046	256	256	1046	256	256	1046
Learning rate	0.0001	0.0001	0.001	0.0001	0.0001	0.001	0.0001	0.0001	0.001

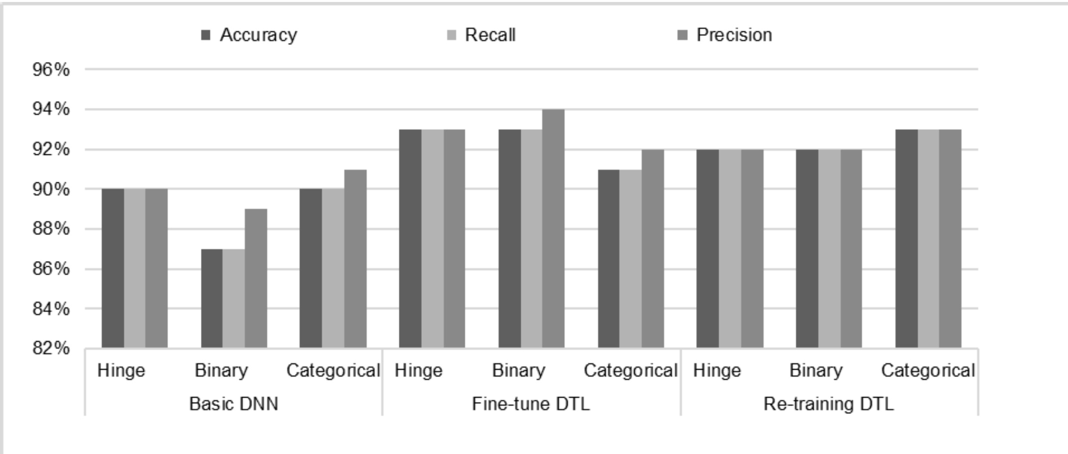

FIGURE 6. Results of DNN classifiers with and without TL (retraining and fine-tuning approaches) using different cross-validation functions.

three different loss functions: the categorical cross-entropy, binary cross-entropy, and hinge functions.

The DTL classifier resulted in a higher accuracy with all loss functions. Although the DNN classifiers achieved a high accuracy of greater than 87%, DTL achieved the highest accuracy result of 93% with the lowest loss rate of 18% with the hinge function. The results of DTL in both TL approaches were similar to those of the binary cross-entropy and hinge functions, but the retraining approach outperformed the fine-tuning approach in terms of accuracy with the categorical cross-entropy function. Figure 6 shows that the highest precision was achieved at 94% with fine-tuning DTL and the binary cross function.

To ensure that the DNN with TL does not have an overfitting problem, we show the loss per epoch for each cross-entropy function and each DNN approach in Figure 5 The loss rate reached approximately 50% with the DNN without TL. The best learning curve was achieved with the DNN with the retraining approach of TL, where the average loss per epoch with the categorical, binary, and hinge functions reached the values of 0.18, 0.24, and 0.22, respectively, as shown in Figs. (5a), (5b), and (5c).

B. DTL AND TRADITIONAL CLASSIFIERS

To evaluate the performance of the classification algorithms, we used different cross-validation methods, namely, holdout (train/test splitting), k-folds cross validation, and leave-one out cross validation (LOOCV). For comparison, we chose the best DTL approach with fine-tuning TL based on the accuracy results in Table 2.

Since we preprocessed the data into a structured, numerical, and normalized format without any missing values, RF achieved good classification results, similar to deep learning. Moreover, RF is less computationally expensive and less prone to overfitting. Deep learning can outperform RF in relatively large datasets with complex data format such as image and speech recognition. The training of the DNN is time consuming and computationally intensive. To find the best model, several variants and combinations of hyperparameters were calculated and evaluated. RF does not require much planning and shows faster training and optimization of hyperparameters. Therefore, the computational cost and time of training RF are relatively small. For the evaluation of the performance in terms of the computational overhead, the time complexity of the DNN is polynomial, $O(n^2)$, whereas RF

TABLE 3. Results of preference recognition using holdout validation and different classifiers: DTL, SVM, RF, and KNN.

Classifier / Metrics	Fine-tuning DTL			SVM	RF	KNN		
	Hinge	Categorical cross	Binary cross			K = 5	K = 3	K = 1
Accuracy	94%	91%	93%	62%	92%	73%	79%	88%
Recall	94%	91%	93%	62%	92%	73%	79%	88%
Precision	94%	92%	94%	64%	93%	75%	81%	90%

has quasilinear complexity $O(n.\log|n|)$. By contrast, RF can reach limited performance with a certain quantity of data, whereas DNN usually benefit from a larger quantity of data and continuously improves the accuracy. To summarize, RF is more robust, requires less computation, and can achieve higher accuracy on our structured data.

In LOOCV, RF reached the best accuracy results at 90% while DTL reached results similar to those of RF at 93% in holdout validation. In k-fold validation, KNN achieved the best accuracy results at 90% and 91% when k was set to 10 and 20, respectively. Because the best accuracy results were achieved using holdout validation, this method was chosen as the base validation for comparison and for tuning the hyperparameter (loss function).

The proposed DTL model was compared with three conventional classification algorithms for EEG signals, namely, SVM, RF, and KNN. Table 3 shows the accuracy, recall, and precision results for RF, KNN, and DTL using the three different loss functions of the categorical cross-entropy, binary cross-entropy, and hinge functions. The KNN classifier led to a better accuracy of 88% when K was set to one.

Although RF achieved a high accuracy of 92%, DTL reached the highest accuracy result of 93% with the hinge function compared to the other conventional classification algorithms. This because the hinge activation function layer at the end of the deep network acts very similarly to the SVM function, i.e., ‘‘maximum-margin’’ classification, to reduce the margin-based loss. Combining the DNN with SVM was achieved by adding the hinge function in the last layer, making it act as a linear-kernel SVM classifier with max-margin loss (L2 regularization).

VIII. CONCLUSION

In this paper, we proposed different DTL model approaches for detecting preferences from EEG signals using the preprocessed DEAP dataset. Two types of features were extracted from the EEG, namely, the PSD and valence. This aspect resulted in a group of 2,367 unique features illustrating the EEG activity in each trial. We used different evaluations measures (accuracy, recall, and precision) and various validation methods (holdout, LOOCV, and k-fold cross validation) to test the classifiers' performances.

We built three DNN models, (1) the basic DNN without TL, (2) fine-tuning of DTL, and (3) retraining of DTL, which achieved accuracies of 87%, 93%, and 92%, respectively. Moreover, we built three different traditional classifiers, namely the RF, SVM, and KNN classifiers, which achieved accuracies of 92%, 62%, and 88%, respectively. These results demonstrate that although the proposed DTL approaches showed higher accuracy, recall, and precision values compared with the KNN and SVM classifiers, RF achieved results similar to those of DTL on the same dataset. We also found that TL improves the performance of the DNN.

We faced challenges associated with choosing the dataset, as there is no benchmark dataset for EEG-based preferences. The dataset was unlabeled, so we detected the preference values by utilizing the knowledge of the emotion domain.

IX. FUTURE WORK

Three main opportunities for future work can be developed: we can consider different neuroimaging techniques for recording brain activity, build and validate alternative feature extraction and classification approaches, and use other TL forms.

First, different neuroimaging can record different brain signals such as MEG, fNIRS, fMRI, and PET. This could extend the knowledge represented in feature spaces rather than recording EEG alone. In addition, our approach can be viewed as a framework of signal classification appropriate for other types of signals, such as ECG. Furthermore, different types of EEG signals - event-related potentials and rhythms (frequency bands)- can be combined as source domain. For example, extracting N200, P300 and alpha frequency can be used to explore the possibilities of applying DTL with different combination of EEG signals. Currently, TL techniques were only applied to one dataset. It is interested to replicate this research on further datasets with other tasks.

Second, developing classification or feature selection algorithms can improve the overall performance for preference detection. The training of the proposed DTL is time-consuming. Therefore, the future works can explore different parameter tuning or even different architecture of deep learning to accelerate the training process. Moreover, it is desirable to apply TL with other conventional classifiers or combining classifiers (such as DTL with RF) using voting, boosting, or stacking algorithms.

Last, we recommend developing different TL techniques. Most TL research focused on cross-subject-to-subject or session-to-session transfers. Devise-to-device transfers have begun to attract attention, but task-to-task transfers remain mostly unexplored. Combining devise-to-device and task-to-task transfers would make EEG classification much more accurate with TL.

ACKNOWLEDGMENT

The authors would like to thank the Deanship of scientific research (DSR) in King Saud University for funding and supporting this research through the initiative of DSR Graduate Students Research Support (GSR).

REFERENCES

[1] X. Gu, Z. Cao, A. Jolfaei, P. Xu, D. Wu, T.-P. Jung, and C.-T. Lin, "EEG-based brain-computer interfaces (BCIs): A survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications," 2020, *arXiv:2001.11337*. [Online]. Available: http://arxiv.org/abs/2001.11337

[2] J. M. Williams, "Deep learning and transfer learning in the classification of EEG signals," Ph.D. dissertation, University of Nebraska, Lincoln, NE, USA, 2017. [Online]. Available: http://digitalcommons.unl.edu/computerscidiss/134

[3] D. Wu, Y. Xu, and B.-L. Lu, "Transfer learning for EEG-based brain-computer interfaces: A review of progress made since 2016," 2020, *arXiv:2004.06286*. [Online]. Available: http://arxiv.org/abs/2004.06286

[4] 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.

[5] R. A. Ramadan, S. Refat, M. A. Elshahed, and R. A. Ali, *Brain-Computer Interfaces* (Intelligent Systems Reference Library), vol. 74, A. E. Hassanien and A. T. Azar, Eds. Cham, Switzerland: Springer, 2015. [Online]. Available: http://link.springer.com/10.1007/978-3-319-10978-7_2 and http://link.springer.com/10.1007/978-3-319-10978-7

[6] J. Teo, L. Hou, J. Tian, and J. Mountstephens, "Classification of affective states via EEG and deep learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 5, pp. 132–142, 2018.

[7] A. Al-Nafjan, M. Hosny, Y. Al-Ohali, and A. Al-Wabil, "Review and classification of emotion recognition based on EEG brain-computer interface system research: A systematic review," *Appl. Sci.*, vol. 7, no. 12, p. 1239, Dec. 2017. [Online]. Available: http://www.mdpi.com/2076-3417/7/12/1239

[8] T. Z. Ramsøy, M. Friis-Olivarius, C. Jacobsen, S. B. Jensen, and M. Skov, "Effects of perceptual uncertainty on arousal and preference across different visual domains," *J. Neurosci., Psychol., Econ.*, vol. 5, no. 4, pp. 212–226, 2012.

[9] M.-H. Lin, S. N. N. Cross, W. J. Jones, and T. L. Childers, "Applying EEG in consumer neuroscience," *Eur. J. Marketing*, vol. 52, nos. 1–2, pp. 66–91, Feb. 2018.

[10] M. Yadava, P. Kumar, R. Saini, P. P. Roy, and D. P. Dogra, "Analysis of EEG signals and its application to neuromarketing," *Multimedia Tools Appl.*, vol. 76, no. 18, pp. 19087–19111, Sep. 2017.

[11] M. A. S. Boksem and A. Smidts, "Brain responses to movie trailers predict individual preferences for movies and their population-wide commercial success," *J. Marketing Res.*, vol. 52, no. 4, pp. 482–492, Aug. 2015.

[12] L. Alvino, E. Constantinides, and M. Franco, "Towards a better understanding of consumer behavior: Marginal utility as a parameter in neuromarketing research," *Int. J. Marketing Stud.*, vol. 10, no. 1, p. 90, Feb. 2018.

[13] R. N. Khushaba, L. Greenacre, S. Kodagoda, J. Louviere, S. Burke, and G. Dissanayake, "Choice modeling and the brain: A study on the electroencephalogram (EEG) of preferences," *Expert Syst. Appl.*, vol. 39, no. 16, pp. 12378–12388, 2012, doi: 10.1016/j.eswa.2012.04.084.

[14] Q. Ma, K. Wang, X. Wang, C. Wang, and L. Wang, "The influence of negative emotion on brand extension as reflected by the change of n2: A preliminary study," *Neurosci. Lett.*, vol. 485, no. 3, pp. 237–240, Nov. 2010.

[15] M. Agarwal and R. Sivakumar, "Cerebro: A wearable solution to detect and track user preferences using brainwaves," in *Proc. 5th ACM Workshop Wearable Syst. Appl.*, 2019, pp. 47–52.

[16] A. Telpaz, R. Webb, and D. J. Levy, "Using EEG to predict consumers' future choices," *J. Marketing Res.*, vol. 52, no. 4, pp. 511–529, Aug. 2015. [Online]. Available: http://journals.ama.org/doi/10.1509/jmr.13.0564

[17] X. Wang, Q. Ma, and C. Wang, "N400 as an index of uncontrolled categorization processing in brand extension," *Neurosci. Lett.*, vol. 525, no. 1, pp. 76–81, Sep. 2012.

[18] J. Wang and W. Han, "The impact of perceived quality on online buying decisions: An event-related potentials perspective," *NeuroReport*, vol. 25, no. 14, pp. 1091–1098, Oct. 2014.

[19] G. Vecchiato, J. Toppi, L. Astolfi, F. De Vico Fallani, F. Cincotti, D. Mattia, F. Bez, and F. Babiloni, "Spectral EEG frontal asymmetries correlate with the experienced pleasantness of TV commercial advertisements," *Med. Biol. Eng. Comput.*, vol. 49, no. 5, pp. 579–583, May 2011.

[20] G. Vecchiato, L. Astolfi, F. D. V. Fallani, F. Cincotti, D. Mattia, S. Salinari, R. Soranzo, and F. Babiloni, "Changes in brain activity during the observation of TV commercials by using EEG, GSR and HR measurements," *Brain Topography*, vol. 23, no. 2, pp. 165–179, Jun. 2010.

[21] G. Cartocci, M. Caratù, E. Modica, A. G. Maglione, D. Rossi, P. Cherubino, and F. Babiloni, "Electroencephalographic, heart rate, and galvanic skin response assessment for an advertising perception study: Application to antismoking public service announcements," *J. Visualized Exp.*, vol. 126, Aug. 2017, Art. no. e55872.

[22] E. Modica, G. Cartocci, D. Rossi, A. C. M. Levy, P. Cherubino, A. G. Maglione, G. Di Flumeri, M. Mancini, M. Montanari, D. Perrotta, P. D. Feo, A. Vozzi, V. Ronca, P. Aricò, and F. Babiloni, "Neurophysiological responses to different product experiences," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–10, Sep. 2018.

[23] R. Ohme, D. Reykowska, D. Wiener, and A. Choromanska, "Analysis of neurophysiological reactions to advertising stimuli by means of EEG and galvanic skin response measures," *J. Neurosci., Psychol., Econ.*, vol. 2, no. 1, pp. 21–31, 2009.

[24] R. Ohme, D. Reykowska, D. Wiener, and A. Choromanska, "Application of frontal EEG asymmetry to advertising research," *J. Econ. Psychol.*, vol. 31, no. 5, pp. 785–793, Oct. 2010.

[25] H. Plassmann, V. Venkatraman, S. Huettel, and C. Yoon, "Consumer neuroscience: Applications, challenges, and possible solutions," *J. Marketing Res.*, vol. 52, no. 4, pp. 427–435, Aug. 2015.

[26] R. Pozharliev, W. J. M. I. Verbeke, and R. P. Bagozzi, "Social consumer neuroscience: Neurophysiological measures of advertising effectiveness in a social context," *J. Advertising*, vol. 46, no. 3, pp. 351–362, Jul. 2017.

[27] B. Touchette and S.-E. Lee, "Measuring neural responses to apparel product attractiveness: An application of frontal asymmetry theory," *Clothing Textiles Res. J.*, vol. 35, no. 1, pp. 3–15, Jan. 2017.

[28] I. Wichakam and P. Vateekul, "An evaluation of feature extraction in EEG-based emotion prediction with support vector machines," in *Proc. 11th Int. Joint Conf. Comput. Sci. Softw. Eng. (JCSSE)*, May 2014, pp. 106–110. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6841851

[29] B. Kaur, D. Singh, and P. P. Roy, "EEG based emotion classification mechanism in BCI," *Procedia Comput. Sci.*, vol. 132, pp. 752–758, 2018.

[30] A. Al-Nafjan, M. Hosny, A. Al-Wabil, and Y. Al-Ohali, "Classification of human emotions from electroencephalogram (EEG) signal using deep neural network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 9, pp. 419–425, 2017.

[31] C. Cohrdes, C. Wrzus, S. Frisch, and M. Riediger, "Tune yourself in: Valence and arousal preferences in music-listening choices from adolescence to old age," *Develop. Psychol.*, vol. 53, no. 9, pp. 1777–1794, Sep. 2017.

[32] I. Michael, T. Ramsoy, M. Stephens, and F. Kotsi, "A study of unconscious emotional and cognitive responses to tourism images using a neuroscience method," *J. Islamic Marketing*, vol. 10, no. 2, pp. 543–564, Jun. 2019.

[33] O. Spray, O. Spray, C. O. O. Ken, A. Worldwide, O. Spray, N. B. C. Today, A. Tour, and L. Angeles, "Designing and managing integrated marketing modern marketing calls for more than developing a good product, pricing it," in *Marketing Management*. Upper Saddle River, NJ, USA: Prentice-Hall, 2016, ch. 17, pp. 475–501.

[34] A. M. Azab, J. Toth, and L. S. Mihaylova, "Signal processing and machine learning for brain-machine interfaces," in *Signal Processing and Machine Learning for Brain-Machine Interfaces*, T. Tanaka and M. Arvaneh, Eds. London, U.K.: Institution Engineering Technology, Sep. 2018, ch. 5. [Online]. Available: https://digital-library.theiet.org/content/books/ce/pbce114e

[35] P. Wang, J. Lu, B. Zhang, and Z. Tang, "A review on transfer learning for brain-computer interface classification," in *Proc. 5th Int. Conf. Inf. Sci. Technol. (ICIST)*, Apr. 2015, pp. 315–322.

[36] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *Proc. Int. Conf. Artif. Neural Netw.* (Lecture Notes in Computer Science), vol. 11141. Rhodes, Greece: Springer, 2018, pp. 270–279. [Online]. Available: http://link.springer.com/10.1007/978-3-030-01424-7_27

[37] G. Xu, X. Shen, S. Chen, Y. Zong, C. Zhang, H. Yue, M. Liu, F. Chen, and W. Che, "A deep transfer convolutional neural network framework for EEG signal classification," *IEEE Access*, vol. 7, pp. 112767–112776, 2019.

[38] S. Raghu, N. Sriraam, Y. Temel, S. V. Rao, and P. L. Kubben, "EEG based multi-class seizure type classification using convolutional neural network and transfer learning," *Neural Netw.*, vol. 124, pp. 202–212, Apr. 2020, doi: 10.1016/j.neunet.2020.01.017.

[39] C. Tan, F. Sun, and W. Zhang, "Deep transfer learning for EEG-based brain computer interface," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, vol. 63, no. 4, pp. 916–920. [Online]. Available: http://arxiv.org/abs/1808.01752 and https://ieeexplore.ieee.org/document/8462115/

[40] A. Uran, C. van Gemeren, R. van Diepen, R. Chavarriaga, and J. D. R. Millán, "Applying transfer learning to deep learned models for EEG analysis," 2019, *arXiv:1907.01332*. [Online]. Available: http://arxiv.org/abs/1907.01332

[41] C. Mühl, S. Koelstra, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan./Mar. 2012.

[42] Z. Lan, Y. Liu, O. Sourina, and L. Wang, "Real-time EEG-based user's valence monitoring," in *Proc. 10th Int. Conf. Inf., Commun. Signal Process. (ICICS)*, Dec. 2015, pp. 1–5. [Online]. Available: http://ieeexplore.ieee.org/document/7459815/

[43] S. D. You and C.-W. Liu, "Classification of user preference for music videos based on EEG recordings," in *Proc. IEEE 2nd Global Conf. Life Sci. Technol. (LifeTech)*, Mar. 2020, pp. 1–2.

[44] S. Shin, S. J. Jang, D. Lee, U. Park, and J. H. Kim, "Brainwave-based mood classification using regularized common spatial pattern filter," *KSII Trans. Internet Inf. Syst.*, vol. 10, no. 2, pp. 807–824, 2016.

[45] M. Aldayel, M. Ykhlef, and A. Al-Nafjan, "Deep learning for EEG-based preference classification in neuromarketing," *Appl. Sci.*, vol. 10, no. 4, pp. 1–23, 2020.

MASHAEL S. ALDAYEL (Member, IEEE) received the B.S. and M.S. degrees in information technology and information system from King Saud University (KSU), Riyadh, Saudi Arabia, in 2010 and 2013, respectively, where she is currently pursuing the Ph.D. degree with the Department of Information Systems. She is also a Lecturer with the Department of Information Technology, KSU. Her current research interests include data science, machine learning, and brain–computer interface. She is a member of ACM and a Co-Founding Member of the KACST ACM SIGCHI Chapter.

MOURAD YKHLEF received the B.Eng. degree in computer science from Constantine University, the M.Sc. degree in artificial intelligence from Paris 13 University, France, and the Ph.D. degree in computer science from the University of Bordeaux 1, France. He is currently a Professor with the Department of Information Systems, King Saud University (KSU), Riyadh, Saudi Arabia. His main research interests include data mining and bioinspired computing.

ABEER N. AL-NAFJAN received the Ph.D. degree in computer science from King Saud University (KSU). She is currently an Assistant Professor with the Department of Computer Science, Al-Imam Mohammed Ibn Saud Islamic University, Riyadh, Saudi Arabia. She is also a Co-Founder with the Human-Computer Interaction (HCI) Design Laboratory. Her research interests include affective computing with special interests in assistive technology and brain–computer interfaces, designing interactive software systems to support people with disabilities in education and communication, and rehabilitation and therapy programs.