



Article Deep Learning Techniques for Pattern Recognition in EEG Audio Signal-Processing-Based Eye-Closed and Eye-Open Cases

Firas Husham Almukhtar ¹, Asmaa Abbas Ajwad ², Amna Shibib Kamil ³, Refed Adnan Jaleel ⁴,*¹, Raya Adil Kamil ³ and Sarah Jalal Mosa ⁵

- ¹ Information Technology Department, Catholic University in Erbil-KRG, Erbil 44003, Iraq
- ² Department of Physiology and Medical Physics, College of Medicine, University of Diyala, Diyala 32001, Iraq
- ³ Department of Medical Device Engineering, Al-Turath University College, Baghdad 61004, Iraq
- ⁴ Department of Information and Communication Engineering, Al-Nahrain University, Baghdad 10071, Iraq
- ⁵ College of Engineering Techniques, Al-Farahidi University, Baghdad 10011, Iraq
- * Correspondence: iraq_it_2010@yahoo.com

Abstract: Recently, pattern recognition in audio signal processing using electroencephalography (EEG) has attracted significant attention. Changes in eye cases (open or closed) are reflected in distinct patterns in EEG data, gathered across a range of cases and actions. Therefore, the accuracy of extracting other information from these signals depends significantly on the prediction of the eye case during the acquisition of EEG signals. In this paper, we use deep learning vector quantization (DLVQ), and feedforward artificial neural network (F-FANN) techniques to recognize the case of the eye. The DLVQ is superior to traditional VQ in classification issues due to its ability to learn a code-constrained codebook. On initialization by the k-means VQ approach, the DLVQ shows very promising performance when tested on an EEG-audio information retrieval task, while F-FANN classifies EEG-audio signals of eye state as open or closed. The DLVQ model achieves higher classification accuracy, higher F score, precision, and recall, as well as superior classification abilities as compared to the F-FANN.

Keywords: signal processing; information retrieval; deep learning vector quantization; feedforward artificial neural network; electroencephalography; classification

1. Introduction

The human brain is made up of an enormous number of nerve cells that interact with one another in a sophisticated network using electrical signals. These signals are transmitted to the cell body after passing through electrochemical solutions in order to modify their impact on that cell's output signal. The electrical signal's current, which travels through these cells, modifies the electrical polarity at the connections between the cell body and dendrites, in which the electrochemical solution can be found [1]. The output of the cell is controlled by the electrochemical junctions' conductivity and updates its state in response to inputs, which is determined by both conscious decision making and environmental feedback [2]. As a result, a lot of focus has been placed on examining the electrical impulses that are sent and received by various brain regions during various actions and states. Despite the availability of various techniques for tracking brain activity in response to specific state, one of the most common practices is the use of EEG, which involves keeping tabs on the brain's electrical signals [3]. This method is gaining popularity, as it is inexpensive and compact equipment is required to gather the signals in comparison to the other methods, such as functional magnetic resonance imaging (fMRI); this is dependent on monitoring variations in cerebral blood flow during certain states or the performance of various tasks [4]. As a result, in recent years, a greater emphasis has been placed on the analysis of EEG data in order to identify these activities and states, plus the detection of anomalies in these signals for use in medicine [5]. A variety of applications



Citation: Husham Almukhtar, F.; Abbas Ajwad, A.; Kamil, A.S.; Jaleel, R.A.; Adil Kamil, R.; Jalal Mosa, S. Deep Learning Techniques for Pattern Recognition in EEG Audio Signal-Processing-Based Eye-Closed and Eye-Open Cases. *Electronics* **2022**, *11*, 4029. https://doi.org/10.3390/ electronics11234029

Academic Editors: Yanhui Guo, Deepika Koundal and Rashid Amin

Received: 22 October 2022 Accepted: 2 December 2022 Published: 5 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). and services have used the analysis of EEG signals, such as medical diagnoses and the brain–computer interface (BCI). BCI enables users of computers to operate without any direct physical contact, where EEG data are examined in order to determine the computer task that will be needed [6]. EEG signals are frequently used with anomaly detection to forecast various diseases. However, in addition to the advantages of understanding the subject's condition at the moment, the EEG data gathered at the same states and actions have shown various EEG patterns in various states of the eye [7]. Eyes often exist in one of two states, which are either closed, known as eye closed (EC) or open, known as eye open (EO). This conduct demonstrates the significance of determining the eye's condition before performing any additional EEG data analysis [8]. Recent years have seen remarkable progress in the pattern-detecting power of DL methods, which have been applied to an ever-increasing amount of input data and analyze the relations between these inputs in order to extract the required knowledge from these data [9].

Processing could be greatly simplified by DL via autonomous end-to-end learning of feature extraction, classification, and preprocessing methods to achieve above-average results on the intended task. In fact, DL architectures have had great success in processing difficult data over the past few years such as audio signals, images, and text, resulting in top-tier results on a wide range of publicly available measures, such as the challenge of visual identification on a broad scale and its expansion in industrial applications [10].

DL is a topic of machine learning that depends on how computational techniques can learn hierarchical representations of incoming data during successive nonlinear transformations. As a result of research into the perceptron and other prior models, deep neural networks (DNNs) have been developed, where (1) a network of artificial "neurons" performs a linear change on the information it receives at each successive layer and (2) a nonlinear activation function is given the linear transformation output by each layer. It is significant that these transformations' parameters are directly found out by systematically reducing a cost function. Despite the widespread use of the term "deep" that refers to neural networks, there is no agreement on what exactly constitutes a "deep" network, and subsequently, what actually qualifies as a deep network and what does not [11].

Acquiring and labeling EEG signals is necessary because DL requires a labeled dataset to draw relevant conclusions and build a model that can be used to predict labels for new inputs. The quality of the predictions made by each classifier must also be evaluated because their abilities to extract accurate knowledge may cause them to perform differently. While unlabeled data are what the predictions are designed for, labeled data are still necessary for measuring their accuracy, so that for the occurrences in the evaluation dataset, the predicted labels are contrasted with the actual labels, which are defined as the testing data. The closer the predicted labels to the actual labels of the test data instances, the better the knowledge extraction, which indicates the better performance of the classifier [12].

2. Literature Survey

The procedure for processing and analyzing EEG data typically consists of two steps: extraction of feature and recognition of patterns [13,14]. Before the popularity of deep learning, the most common way to extract features was to use signal analysis to pull out time–frequency features, such as spectral density, density of power [15], band power [16], separate components [17], and differential entropy [18]. The widely researched recognition on pattern and machine learning techniques contains ANN [19,20], naive Bayes [21], support vector machines [22,23], etc. Due to deep advertising and widespread use of DL, an ever-increasing number of neuroscience and brain study teams are discovering its strength in building techniques to use EEGs to intelligently comprehend and analyze brain activity, hence offering an end-to-end approach that unifies the extraction of features, classification, and clustering. To categorize mental stresses, the authors of [24] developed a multichannel DNN. In [25], an LSTM network was used to categorize forms of motor imagery and extracted the network's useful characteristics using a one-dimensional aggregation approximation technique. In [26], the authors employed a CNN-based predictive

modeling strategy for estimating the ages of brains. Their research revealed that the brain's ability to estimate age is extremely accurate. With their suggested spatiotemporal deep convolution model, the authors of [27] significantly enhanced the accuracy of driver fatigue detection by highlighting the significance of geographical data and the timing of EEGs. To automatically detect epileptic convulsions in EEG recordings, further, a full-stack multiview DL architecture was suggested in [28].

In [29], the authors attempted to build CNN with transfer learning in mind and successfully used the model to diagnose mild depression in patients. Furthermore, a mixed neural network of LSTM on domain of time–frequency data was trained using an activation function of rectified linear unit (ReLU) to classify sleep stages [30]. Later on, a compact model based on a complete convolutional network (EEGNet) was presented for EEG for the various BCI classification tasks [31]. The authors in [32] suggested a parallel and cascaded convolution recurrent neural network technique by efficient learning of spatiotemporal representation to distinguish the human motion commands of actual EEG signals. Moreover, in [33], EEG data are transformed into EEG-dependent optical flow and video information, which is classified by RNN and CNN, for the development of a successful BCI-based rehabilitation support system.

Large amounts of content information and varied visual qualities are packed into multimedia data that are commonly employed in the collection and analysis of EEG data [34–36]. Through the study of EEG signals, researchers attempted to determine and categorize the content information of users' watched multimedia material [37–39].

The authors of [35] created a mapping link between natural image attributes and EEG representation using LSTM network learning to develop a model for EEG responses to visual cues. The improved EEG signal representation was then used for the classification of natural images. These DL-based strategies produced remarkable classification results, especially when compared to conventional approaches.

Moreover, twenty-eight university students' EEG data were gathered while they were resting with their eyes closed and their eyes open; then, Fourier transform was applied in the theta, delta, beta, and alpha bands and analyzed in nine regions across the scalp to make estimations of total power [40]. Arousal level was also determined by measuring skin conductance. Topographic effects of the situation are shown in Figure 1.

Resting EEG data from 70 subjects were analyzed by the authors of [41] (46 adults, 29 male), across testing sessions spaced by 12 ± 1.1 years. Alpha of EEG was separated, quantified, and identified by applying reference-free techniques that combine principal component analysis (PCA) with current source density (CSD). Measures of overall (EC-plus-EO) and net (EC-minus-EO) posterior amplitude of alpha and inconsistencies in asymmetry were compared among several trials. Waves 4 and 6 of the resting EEG's CSD-fPCA structure are shown in Figures 2 and 3 at 13 electrode sites common to both waves and topographies of the mean alpha of factor scores, respectively.

Recent research has demonstrated that by mining EEG data, multimedia content information may be reconstructed. In [34], the authors developed a strategy for deducing the information of visual inputs from electrical brain activity. By applying generative adversarial networks (GANs) and a variable-valued auto-encoder (VAE), they discovered that patterns relating to visual content can be observed in EEG data, and images that are semantically compatible with the incoming visual stimuli can be produced using the content. Despite the fact that these techniques have proven that a DL framework may be used for EEG-based image classification, the input is frequently the actual EEG measurements or time–frequency properties retrieved using signal analysis methods, and certain aspects of human brains have not been given much thought, such as hemispheric lateralization, and the accuracy of classification produced is 82.9% [35].



Figure 1. Environmental influences on topography (**a**) Delta (**b**) Theta (**c**) Alpha (**d**) Beta. One EEG band is active in each row. In the left column, we see the typical power output across all conditions. Power distributions for the EO and EC circumstances are displayed in the middle columns; the power decrease in each band when the eyelids open can be easily observed. The rightmost column displays standard scores representing the difference (EO-EC) at each electrode site; note the focal changes apparent in theta, beta, and delta bands.



Figure 2. Waves 4 and 6 of the resting EEG's CSD-fPCA structure with alpha prefiltering at 13 electrode locations that are shared by both waves. (**A**): Separation of low-frequency alpha signals can be seen in factor loadings (9.37 Hz peak; blue) and alpha of high frequency (10.15 Hz peak; red) from a broad delta factor (1.56 Hz peak, green). (**B**): Topographies of mean factor scores for EO and EC condition (averaged of waves 6 and 4).



Figure 3. Mean factor score alpha topographies (pooled across low- and high-frequency alpha) for top row (EO), middle row (EC), and overall alpha (EO + EC; bottom row), independently offered for each wave.

3. Materials and Methods

Many applications and analyses that use EEG signals require knowledge of the subject's eye state, whether it is open or closed. DLVQ and F-FANN techniques can be used to predict a class, or label, per each data instance, depending on the attributed values of that instance, which makes them applicable for predicting the state of the subject's eyes, based on the values collected using EEG. However, in order to use a classifier in a certain application, it is important to train that classifier using data collected from the same environment, where each instance is labeled according to the actual state of that instance. Thus, to use DLVQ and F-FANN techniques in predicting the state of the eye based on the EEG signals, data must be collected from different subjects in a controlled environment, i.e., the actual state of their eyes are logged alongside with the EEG signals. However, DLVQ and F-FANN techniques may have different performance depending on the input data, which imposes the need to evaluate the performance of the classifiers, in order to select the most appropriate one. The rationale behind the comparison of DLVQ and F-FANN in EEG signal processing is that we can demonstrate that the activations and parameters of a neural network can be quantized using product quantization with shared subdictionaries without materially affecting the network's accuracy.

3.1. Data Collection and Classification

In order to train classifiers and evaluate their performances, a labeled dataset is required so that classifiers extract the relations between the attribute values of each instance and the label given to it, while evaluation is conducted by comparing the predictions provided by the classifiers to the actual labels of the instances used in the evaluation. For this purpose, the EEG data included 150 recordings from 27 participants, taking around 24 s in each condition. EEG signals are collected from a 16-channel V-AMP amplifier at 1024 samples/second, sample rate, while the impedance of the channels is maintained below 5 k Ω . The electrodes that collect the EEG signals are positioned at (Fp2, Fp1, Fz, F3, FCz, F4, T3, T4, CPz, Cz, Pz, P7, C3, P8, C4, and Oz). Undesired frequencies are filtered out

using three types of filters, an 80 Hz online low-pass filter, a 0.1 Hz high-pass filter, and a 50 Hz notch filter. Moreover, two reference nodes are connected to the subject's ears, one to each ear. After preprocessing the signals, using the Brain Vision Analyzer, the collected data are referenced using the average EEG of all the collected channels, per each instance, and the sampling rate is reduced to 256 Hz. Artifact segments and epochs with amplitudes greater than 150 μ V are marked for removal for further analysis. Finally, the band power values of the alpha, beta, delta, and theta are computed from the collected EEG signals, which results in 64 attributes that describe each data instance in the collected dataset. When all of the training data are utilized at once, this is called an epoch, and it is measured in terms of the total number of training data iterations in a single cycle.

The performance of DLVQ and F-FANN is evaluated in this study in order to select the classifier with the best performance for EEG signal classification, to forecast the state of the subject's eye, depending on the data extracted from the EEG signals. The data collected from EEG signals are used to train the DL techniques and evaluate their performance, which reflects the quality of the knowledge extracted from the training data.

3.2. Deep Learning Vector Quantizer

It is possible to obtain a frame-level codeword and initial codebook sequence using the level structured k-means using VQ approach described in [42]. These data can then be used for DLVQ. DLVQ is based on the LVQ principle, which has been found to be helpful in several disciplines, including ASR and text classification, and uses the power of deep learning simultaneously. However, the difference between this method and the method given in [42] is that in this study, it is applied on brain signals and not on heart signals, and comparison is carried out between DLVQ and F-FANN on EEG signals.

3.2.1. DLVQ System Structure

Similar to the DLVQ approach used in [42], which employs DNN as a code-book learner and VQ, this study follows the same basic outline. As with DNN-based ASR, a DNN can be trained using the frame-level label information provided from the initial quantizer. Figure 2 depicts the overarching framework of the training program. K-means is used to first learn an initial codebook using training frames (No enclosing contexts are employed). Then, using normal VQ, each frame's codeword is acquired. Finally, a DNN is trained using optimization objective for cross-entropy, with the codeword serving as the class target for every individual frame.

3.2.2. DNN Training

A core frame was spliced into the DNN's input (whose label is the splice label) and its left and right sides have n context frames, e.g., n = 6 or n = 8. The sigmoid units were used to build the hidden layers, and a softmax layer was used for the output. It has exactly the same number of nodes as the VQ initializer's codeword. DNN's fundamental structure is depicted in Figure 4. In particular, an expression for the node values looks like Equations (1) and (2),

$$\mathbf{v}\mathbf{x}^{i}\mathbf{\&} = \begin{cases} Wight_{1}\mathbf{o}^{t} + \mathbf{b}\mathbf{v}_{1}, & i = 1\\ Wight_{i}\mathbf{y}^{i} + \mathbf{b}\mathbf{v}_{i}, & i > 1 \end{cases}$$
(1)

$$\mathbf{v}\mathbf{y}^{i} \& = \begin{cases} \text{sigmoid}(\mathbf{x}^{i}), & i < tn\\ \text{softmax}(\mathbf{x}^{i}), & i = tn \end{cases}$$
(2)

where \mathbf{bv}_1 , \mathbf{bv}_i are the bias vectors; *tn* is the softmax and sigmoid functions; and the total number of hidden layers are element-wise operations. The vector \mathbf{vx}^i corresponds to activations of prenonlinearity and \mathbf{vy}^i is the vector of neuron at the *i*th hidden layer. Codeword posterior estimates were derived from the softmax outputs as in Equation (3):

$$P(CW_j \mid \mathbf{o}_t) = \mathbf{E}_t^n(j) = \frac{\exp(\mathbf{x}_t^n(j))}{\sum_i \exp(\mathbf{x}_t^n(i))}$$
(3)



where CW_i represents the *j*th codeword and $\mathbf{E}_t^n(j)$ is the *j*th element of $\mathbf{E}_t^n(j)$.

Figure 4. Structure of DLVQ system.

Through increasing the log posterior probability across the training frames, DNN was trained. This is the same as trying to minimize the loss function with the largest negative cross-entropy. Let X represent the entire training set with N frames, i.e., $x_{1:N}^0 \in X$, then the loss with respect to X is given by Equation (4):

$$\mathcal{L}_{1:N} = -\sum_{t=1}^{N} \sum_{j=1}^{J} \mathbf{l}_{t}(j) \log P(CW_{j} \mid \mathbf{o}_{t})$$
(4)

where $P(CW_j | \mathbf{o}_t)$ is mentioned in Equation (3); \mathbf{l}_t is the vector of label at frame t, which is the pseudo one obtained from the initializer of k-means VQ. Utilizing error backpropagation, we are able to reduce the loss objective function, which is a gradient-descentdependent optimization technique that is advanced for neural networks. Calculating partial derivatives of the function of loss objective in relation to the output layer's prenonlinearity activations Xⁿ will produce the vector of error to be backpropagated to the previous hidden layers. In the previous hidden layer, backpropagated error vectors are described in Equations (5) and (6):

$$\epsilon_t^n = \frac{\partial \mathcal{L}_{1:N}}{\partial \mathbf{x}^n} = \mathbf{E}_t^n - \mathbf{l}_t \tag{5}$$

$$\boldsymbol{\epsilon}_t^i = \boldsymbol{W}_{t+1}^T \boldsymbol{\epsilon}_t^{i+1} * \mathbf{y}^i * \left(1 - \mathbf{E}^i\right), i < n \tag{6}$$

where * refers to element-wise multiplication. Vectors of error from specific hidden layers combined with the overall gradient with respect to the matrix of weight through training W_i are computed by Equation (7):

$$\frac{\partial \mathcal{L}_{1:N}}{\partial W_i} = \mathbf{C}_{1:N}^{i-1} \left(\epsilon_{1:N}^i \right)^T \tag{7}$$

From Equation (7), it is observed that above both $C_{1:N}^{i-1}$ and $\epsilon_{1:N}^{i}$ are measures, which are constructed by stringing together vectors representing each training frame, from frame 1 to frame *N*, i.e., $\epsilon_{1:N}^{i}$ 1: $N = [\epsilon_{1}^{i}, \ldots, \epsilon_{t}^{i}, \ldots, \epsilon_{N}^{i}]$. Parameters are recalculated using the gradient in Equation (7), a batch-based gradient-descent update, only once. Parallelization can thus be readily carried out to hasten the learning process after each sweep across the entire training set. Stochastic gradient descent (SGD), on the other hand, typically functions

more effectively in practice. This is because SGD assumes that the true gradient may be approximated by the gradient at a single frame t, i.e., $\mathbf{C}_1^{i-1}(\boldsymbol{\epsilon}_1^i)^T$, and each frame's parameters are updated immediately after viewing. The minibatch SGD is more popular because all of the matrices fit into the GPU memory due to the minibatches' appropriate size, resulting in a more computationally efficient learning procedure. In this work, the parameters are updated using minibatch SGD.

In order to maximize the accuracy of the DNN, it is best to train it with a cross-entropy loss function that minimizes the likelihood that it will forget any labels it has been given by its initializer of VQ; that is, a "perfect" training cycle will enable the DNN to achieve the same VQ outcomes as its initializer. In contrast, low frame accuracy was reported throughout the realistic training approach: less than half for the testing and training data. This shows that DNN is capturing new information in the input rather than learning exactly what its initializer does.

3.3. Feedforward Artificial Neural Network

F-FANN is implemented to predict the state of the eye, depending on the input values collected from the EEG signals. As each instance consists of a one-dimensional vector, with 64 values, the implemented neural network uses only fully connected layers. According to the number of attributes in the data, the number of neurons in the input layer is set to 64 neurons: 1 neuron per each input value. This input layer is linked to the first hidden layer, which consists of 256 neurons. In addition to this hidden layer, 3 more hidden layers are used before the output layer, with 256, 128 and 64 neurons, sequentially, producing a total of 4 hidden layers. The output layer consists of a single neuron, as a single output is required from the neural network to describe the probability of the input to be collected from a subject with EC state. A summary of the implemented feedforward artificial neural network is shown in Figure 5.



Figure 5. F-FANN for EEG classification.

The use of such topology allows for the extraction of complex features, from the input, without dramatically increasing the complexity of the computation in the neural network, which requires more computer resources or execution time. Moreover, according to the benefits of the ReLU activation function, including the faster learning and elimination of the vanishing gradient problem, all hidden layers use this activation function. As the output required from the neural network is limited to the range from zero to one, this neuron uses the Sigmoid activation function. In artificial neural networks, overfitting occurs when the neural network is emphasized on a certain path, among neurons, to reach the required solution. As the training continues in an overfitted neural network, more emphasis is added to that batch, by amplifying the weights among neurons in that path, during backpropagation. Thus, to avoid such behavior, a predefined percentage of the neurons in every layer are dropped through the training, which is defined as the dropout rate. These neurons are selected randomly per each training epoch, so that the neural network is enforced to find multiple paths to come up with the same prediction. Relying on different features, enforced by the dropout, the output of the neural network considers all these features, which eliminates the errors that may occur according to the strict reliance on a certain feature. Figure 5 illustrates an example of dropout during training.

4. Performance Evaluation and Results

In order to choose the classifier with the best assessment in EEG signal classification, to predict the eye state of the subject, predictions provided by each DL classifier are compared to the original states of the eyes in the dataset, by distributing these predictions and the actual states in the confusion matrix. The true EO represents the number of instances that are collected from subjects with their eyes open and forecasted by the DL classifier as EO. False EO is the number of instances that have EO labels, while the classifier predicts them as EC. True EC represents the number of EC instances that are correctly predicted by the classifier as EC, while false EC is the number of EO instances that are predicted as EC by the classifier. Using the values in the confusion matrix created based on the classification results of a certain classifier, the measures of performance are applied to describe the performance of that classifier. Thus, the accuracy of the predictions is calculated using Equation (8). Moreover, the precisions of the predictions provided for the EO and EC classes are shown in Equations (9) and (11) sequentially, while the recalls of each of these classes are calculated using Equations (10) and (12). These values are then used to calculate the F Scores for the EO and EC classes, according to Equations (13) and (14). Moreover, as some of the applications that rely on EEG classification to estimate the state of the subject's eye require faster decisions, the average time required by each classifier to produce a prediction for a single instance is also measured. Based on these measures, the DL techniques with the best performance can be selected for the purpose of eye state prediction based on the EEG signals.

$$ACC = \frac{\text{True EC} + \text{True Eo}}{\text{True EC} + \text{False EC} + \text{True EO} + \text{False EO}}$$
(8)

$$Precision_{EC} = \frac{True EC}{True EC + False EC}$$
(9)

$$\operatorname{Recall}_{\mathrm{EC}} = \frac{\operatorname{True} \operatorname{EC}}{\operatorname{True} \operatorname{EC} + \operatorname{False} \operatorname{EO}}$$
(10)

$$Precision_{EO} = \frac{True EO}{True EO + False EC}$$
(11)

$$Recall_{EO} = \frac{True EO}{True EO + False EC}$$
(12)

$$FScore_{EO} = 2 \times \frac{Precision_{EO} \times Recall_{EO}}{Precision_{EO} + Recall_{EO}}$$
(13)

$$F Score_{EC} = 2 \times \frac{Precision_{EC} \times Recall_{EC}}{Precision_{EC} + Recall_{EC}}$$
(14)

Each channel's EEG signal is first adjusted to one variance and zero mean. Additionally, it is segmented into adjacent frames (each one lasts for one second, which is equivalent to the length of one hundred and fifty samples). The result is 10 frames transmitted on each channel. Figure 6 shows an illustration of this procedure.

Three level-structured k-means VQ systems were developed for use as benchmarks. with 128 (8 clusters on first level, and 4 on the second level) 256 (three levels with 16, 4 and 2 clusters in every level), and 512 (4 levels with 32, 8, 16, and 4 clusters in each level) codewords characterized by 128 k-means, 256 k-means, and 512 k-means.

The number of times each codeword appears in an audio EEG clip was used to create the BoW vector representation of that clip. The pseudo codeword labels produced by the systems of baseline k-means were used to build DLVQ systems. The input of all DNNs is a splice of the center frame and its 8 context frames, and each layer of the 7 hidden layers has 2048 nodes as in [42]. Each system's output softmax layer shares the same dimensionality as the codebook vocabulary of its respective VQ initializer system, that is, 128, 256, and 512, respectively. Based on the Kaldi voice recognition tools, we developed the DNN systems. The DNN is trained using the following method: layer-by-layer generative pretraining is used for parameter initialization. As a next step, we use backpropagation and the cross-entropy goal function to train the network discriminatively. The initial learning rate is set at 0.09, and the minibatch size is set to 256. Then, frame accuracy is checked on the development set after every training iteration, the learning rate is reduced by a factor of 0.5% if the improvements are less than 0.5%. After the accuracy of frame enhancement drops to less than 0.1%, the training procedure is terminated. The 128-codeword k-means-based DNN trained in practice obtained 34% frame accuracy on the training and test datasets, respectively; the 256-codeword k-means model obtained 23% and 29%, respectively; the model dependent on 512-codeword k-means obtained 24% and 27%. Figure 5 shows these findings, which show that the frame EEG accuracy-shifting tendencies in the training and development set are similar and primarily growing. This demonstrates that the DNN can successfully imitate its VQ initializer through cross-entropy training (the "labels" by k-means VQ being retained); however, because the accuracy of the last frame was below 50%, we may infer that the DNN is not figuring out the specifics of its initializer's operation, but rather is actively gathering fresh data. The representation of BoW of an audio EEG clip was then made by running the clip's frames through a trained DNN and adding up the resulting vectors. For both the baseline and suggested frameworks, as a histogram, the properties of the BoW vector representation for each clip were normalized so that they added up to 1. HIK kernel and SVMs were applied as the classifiers. It is evident that in MAP, DLVQ achieves a 4.5% relative increase over the k-means baseline. An approximate 10.5% relative gain was obtained when fusing the findings of the baseline and proposed systems. DLVQ picks up some supplementary data that k-means miss. According to their AP scores on the development set, the two systems' classifier scores from the basic late fusion approach are simply weighted together. These encouraging results demonstrate that DLVQ does aid in improving the representative power of VQ-based BoW vectors. Figure 7 offers the accuracy of DLVQ with different codebooks in training sets. Figure 8 offers the accuracy of DLVQ with different codebooks in development sets.



Figure 6. Amplitude versus time for EEG signal.

The F-FANN is implemented using the Keras library and evaluated using the 5-fold cross-validation method. Per each iteration in the cross-validation, the model is trained for 1000 epochs using the training bins and evaluated using the testing bin. These results are used to calculate the performance evaluation measures. The average time consumed by the feedforward artificial neural network to come up with a prediction per data instance is 0.009 ms.

The results show that DLVQ scores better overall accuracy than the feedforward neural network. Higher precision is scored by the feedforward artificial neural network in predicting the EO state, while higher recall is scored in the EC predictions. However, the F score for both states is equal, which is 91%. Thus, the overall F score scored 91% as



well. Figure 9 illustrates the accuracy of F-FANN in training sets, while Figure 10 shows the accuracy of F-FANN in development sets. Tables 1 and 2 also the Figures 11 and 12 offer the precision, recall, and F score for DLVQ and F-FANN, respectively.

ACC of Frame EEG on Training Set with 128 Codebook
 ACC of Frame EEG on Training Set with 512 Codebook



Figure 7. Accuracy of DLVQ with different codebooks in training sets.

ACC of Frame EEG on Dvelopment Set with 128 Codebook
 ACC of Frame EEG on Development Set with 512 Codebook

Figure 8. Accuracy of DLVQ with different codebooks in development sets.



Figure 9. Accuracy of F-FANN with different codebooks in training sets.



____ACC of Frame EEG on Development Set

Figure 10. Accuracy of F-FANN in development sets.

Table 1. Precision, recall, and F score for DLVO.

Class	Precision	Recall	F Score
EO	90	92	91
EC	92	90	91



Table 2. Precision, recall, and F score for F-FANN.

Figure 11. Precision, recall, and F score for DLVO technique.



Figure 12. Precision, recall, and F score for F-FANN technique.

F-FANN shows the highest overall performance measure, with an average prediction time of 0.009 mS per each data instance. Moreover, DLVQ is also able to average a prediction time of 0.074 mS.

5. Limitations and Optimal Points

Large volumes of data are a major roadblock for this proposed work. It can be costly to train it using huge and complicated data models. A lot of hardware is also required to perform complicated mathematical computations. There is no standard or single way to choose DL tools. It is not always possible to obtain answers using DL algorithms when dealing with interdisciplinary issues. With DL, a perfect solution might not always be possible. Inaccurate or incorrect output might result from poor-quality, incomplete, or incorrect data. In fact, DL may not be able to answer issues that are not provided in a classification format, as its methods are optimized for such situations.

The optimal points are briefed as follows: the best features of the proposed system are DLVQ and F-FANN, which perform well with unstructured or unlabeled data, as there are different DL algorithms, libraries, and open-source frameworks available. The number of practical uses for them is extensive, scalable, and efficient.

DLVQ and F-FANN make it easier to automatically recognize features without first extracting those characteristics. One neural network-based technique may be modified and applied to a variety of data kinds and applications, since it is a resilient system.

6. Conclusions and Recommendation

In this paper, EEG signals were collected from 27 participants in order to evaluate the performance of 3 of the DL techniques, namely DLVQ and F-FANN, to predict the state of the subject's eye based on the collected EEG signals. The collected data were split into five bins, where each bin was used once for evaluation while the remaining bins were used for training, using a 5-fold cross-validation evaluation approach. This approach ensures unbiased evaluation, where data instances in a randomly selected testing set may be more suitable for one classifier than another, which produces biased evaluation measures. DLVQ showed the highest overall performance measure. Additionally, we provide a discriminative approach to LVQ in this research, employing a DL framework to extract a superior VQ representation from the initializer baseline VQ systems. When combined with its k-means VQ initializer, the DLVQ system is able to capture novel information and achieve a highly encouraging relative performance improvement.

There are still many areas where DL techniques in training may be enhanced. We would also like to examine DLVQ's and F-FANN's performance in other fields, such as computer vision, and investigate the theoretical relationship between DLVQ and its initializers. Furthermore, the integration of DLVQ and F-FANN with preexisting technology has become more feasible, including the brain–computer interface, big data, and the Internet of things (IoTs).

Author Contributions: Methodology, F.H.A.; conceptualization, A.A.A.; writing—original draft, A.S.K.; review and editing, R.A.K.; software and supervision, R.A.J.; Methodology, S.J.M. All authors have read and agreed to the published version of the manuscript.

Funding: This paper received no external funding.

Data Availability Statement: The data shall be made available on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Zhong, H.; Wang, J.; Li, H.; Tian, J.; Fang, J.; Xu, Y.; Jiao, W.; Li, G. Reorganization of Brain Functional Network during Task Switching before and after Mental Fatigue. *Sensors* 2022, 22, 8036. [CrossRef] [PubMed]
- Jackson, A.F.; Bolger, D.J. The neurophysiological bases of EEG and EEG measurement: A review for the rest of us. *Psychophysiology* 2014, 51, 1061–1071. [CrossRef] [PubMed]
- Perentos, N.; Nicol, A.U.; Martins, A.Q.; Stewart, J.E.; Taylor, P.; Morton, A.J. Techniques for chronic monitoring of brain activity in freely moving sheep using wireless EEG recording. *J. Neurosci. Methods* 2017, 279, 87–100. [CrossRef] [PubMed]
- 4. Asayesh, A.; Ilen, E.; Metsäranta, M.; Vanhatalo, S. Developing Disposable EEG Cap for Infant Recordings at the Neonatal Intensive Care Unit. *Sensors* 2022, 22, 7869. [CrossRef] [PubMed]

- Dahl, G.E.; Yu, D.; Deng, L.; Acero, A. Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition. *IEEE Trans. Audio Speech Lang. Process.* 2011, 20, 30–42. [CrossRef]
- Hinton, G.; Deng, L.; Yu, D.; Dahl, G.E.; Mohamed, A.-R.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.N.; et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Process. Mag.* 2012, 29, 82–97. [CrossRef]
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–8 December 2012; pp. 1106–1114.
- Byun, B.; Kim, I.; Siniscalchi, S.M.; Lee, C.-H. Consumer-level multimedia event detection through unsupervised audio signal modeling. *Interspeech* 2012, 2012. [CrossRef]
- Rishi, S.; Debnath, S.; Dewani, S.; David, D.S.; Jalee, R.A.; Zahra, M.M.A. AI-Based convolute Neural Approach Management To Predict The RNA Structure. In Proceedings of the 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 28–29 April 2022; pp. 2224–2228. [CrossRef]
- Mahdi, R.D.; Qasim, M.A.; Allayla, N.M.; Jaleel, R.A. A Customized Iomt-Cloud Based Healthcare System For Analyzing of Brain Signals Via Supervised Mining Algorithms. *J. Eng. Sci. Technol.* 2022, 76–83. Available online: https://www.researchgate.net/publication/ 359622556_A_customized_IOMT-_cloud_based_healthcare_system_for_analyzing_of_brain_signals_via_supervised_mining_ algorithms (accessed on 1 February 2022).
- Longo, L. Modeling Cognitive Load as a Self-Supervised Brain Rate with Electroencephalography and Deep Learning. *Brain Sci.* 2022, 12, 1416. [CrossRef]
- 12. Kumar, P.; Abubakar, A.A.; Sazili, A.Q.; Kaka, U.; Goh, Y.-M. Application of Electroencephalography in Preslaughter Management: A Review. *Animals* **2022**, *12*, 2857. [CrossRef]
- Gao, Z.; Wang, S. Emotion recognition from EEG signals by zleveraging stimulus videos. In *Proceedings of the 16th Pacific-Rim Conference on Advances in Multimedia Information Processing–PCM 2015*; Part II; Springer: Berlin/Heidelberg, Germany, 2015; pp. 118–127.
- 14. Kim, M.K.; Kim, M.; Oh, E.; Kim, S.P. A review on the computational methods for emotional state estimation from the human EEG. *Comput. Math Methods Med.* **2013**, *5*73734. [CrossRef] [PubMed]
- Lin, Y.-P.; Wang, C.-H.; Jung, T.-P.; Wu, T.-L.; Jeng, S.-K.; Duann, J.-R.; Chen, J.-H. EEG-based emotion recognition in music listening. *IEEE Trans. Biomed. Eng.* 2010, 57, 1798–1806.
- 16. Dahne, S.; Biessmann, F.; Meinecke, F.C.; Mehnert, J.; Fazli, S.; Muller, K.-R. Integration of Multivariate Data Streams with Bandpower Signals. *IEEE Trans. Multimedia* **2013**, *15*, 1001–1013. [CrossRef]
- 17. Cong, F.; Alluri, V.; Nandi, A.K.; Toiviainen, P.; Fa, R.; Abu-Jamous, B.; Gong, L.; Craenen, B.G.W.; Poikonen, H.; Huotilainen, M.; et al. Linking Brain Responses to Naturalistic Music Through Analysis of Ongoing EEG and Stimulus Features. *IEEE Trans. Multimedia* 2013, *15*, 1060–1069. [CrossRef]
- Duan, R.-N.; Zhu, J.-Y.; Lu, B.-L. Differential entropy feature for EEG-based emotion classification. In *Proceedings of the 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER), San Diego, CA, USA, 6–8 November 2013*; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA, 2013; pp. 81–84.
- King, L.; Nguyen, H.T.; Lal, S. Early driver fatigue detection from electroencephalography signals using artificial neural net-works. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2006, 2006, 2187–2190. [PubMed]
- Saha, A.; Konar, A.; Chatterjee, A.; Ralescu, A.; Nagar, A.K. EEG analysis for olfactory perceptual-ability measurement using a recurrent neural classifier. *IEEE Trans. Human Mach. Syst.* 2014, 44, 717–730. [CrossRef]
- Chan, A.; Early, C.E.; Subedi, S.; Li, Y.; Lin, H. Systematic analysis of machine learning algorithms on EEG data for brain state intelligence. In Proceedings of the 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Washington, DC, USA, 9–12 November 2015; pp. 793–799.
- Kawakami, T.; Ogawa, T.; Haseyama, M. Novel image classification based on decision-level fusion of EEG and visual features. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 5874–5878. [CrossRef]
- 23. Zhang, J.; Yin, Z.; Wang, R. Pattern Classification of Instantaneous Cognitive Task-load Through GMM Clustering, Laplacian Eigenmap, and Ensemble SVMs. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2016**, *14*, 947–965. [CrossRef]
- Jiao, Z.; Gao, X.; Wang, Y.; Li, J.; Xu, H. Deep Convolutional Neural Networks for mental load classification based on EEG data. Pattern Recognit. 2017, 76, 582–595. [CrossRef]
- Wang, P.; Jiang, A.; Liu, X.; Shang, J.; Zhang, L. LSTM-Based EEG Classification in Motor Imagery Tasks. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2018, 26, 2086–2095. [CrossRef]
- Cole, J.H.; Poudel, R.P.; Tsagkrasoulis, D.; Caan, M.W.; Steves, C.; Spector, T.D.; Montana, G. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage* 2017, 163, 115–124. [CrossRef]
- Gao, Z.; Wang, X.; Yang, Y.; Mu, C.; Cai, Q.; Dang, W.; Zuo, S. EEG-based spatio-temporal convolutional neural network for driver fatigue evaluation. *IEEE Trans. Neural. Netw. Learn. Syst.* 2019, 30, 2755–2763. [CrossRef] [PubMed]
- Yuan, Y.; Xun, G.; Jia, K.; Zhang, A. A Multi-View Deep Learning Framework for EEG Seizure Detection. *IEEE J. Biomed. Health Inform.* 2018, 23, 83–94. [CrossRef] [PubMed]
- Li, X.; La, R.; Wang, Y.; Niu, J.; Zeng, S.; Sun, S.; Zhu, J. EEG-based mild depression recognition using convolutional neural network. *Med Biol. Eng. Comput.* 2019, 57, 1341–1352. [CrossRef]

- Dong, H.; Supratak, A.; Pan, W.; Wu, C.; Matthews, P.M.; Guo, Y. Mixed neural network approach for temporal sleep stage classification. *IEEE Trans. Neural. Syst. Rehabil. Eng.* 2018, 26, 324–333. [CrossRef]
- 31. Lawhern, V.J.; Solon, A.J.; Waytowich, N.R.; Gordon, S.M.; Hung, C.P.; Lance, B.J. EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces. *J. Neural Eng.* **2018**, *15*, 056013. [CrossRef] [PubMed]
- Zhang, D.; Yao, L.; Zhang, X.; Wang, S.; Chen, W.; Boots, R.; Benatallah, B. Cascade and Parallel Convolutional Recurrent Neural Networks on EEG-based Intention Recognition for Brain Computer Interface. Proc. Conf. AAAI Artif. Intell. 2018, 32. [CrossRef]
- Tan, C.; Sun, F.; Zhang, W.; Chen, J.; Liu, C. Multimodal classification with deep convolutional-recurrent neural networks for electroencephalography. In *Neural Information Processing*; Liu, D., Xie, S., Li, Y., Zhao, D., El-Alfy, E.-S.M., Eds.; Springer: Cham, Switzerland, 2017; pp. 767–776.
- 34. Kavasidis, I.; Palazzo, S.; Spampinato, C.; Giordano, D.; Shah, M. Brain2image: Converting brain signals into images. In *Proceedings* of the 2017 ACM on Multimedia Conference; ACM: New York, NY, USA, 2017; pp. 1809–1817.
- Spampinato, C.; Palazzo, S.; Kavasidis, I.; Giordano, D.; Souly, N.; Shah, M. Deep learning human mind for automated visual classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6809–6817.
- 36. Righart, R.; de Gelder, B. Rapid influence of emotional scenes on encoding of facial expressions: An ERP study. *Soc. Cogn. Affect Neurosci.* 2008, *3*, 270–278. [CrossRef] [PubMed]
- Das, K.; Giesbrecht, B.; Eckstein, M.P. Predicting variations of perceptual performance across individuals from neural activity using pattern classifiers. *NeuroImage* 2010, 51, 1425–1437. [CrossRef]
- Wang, J.; Pohlmeyer, E.; Hanna, B.; Jiang, Y.-G.; Sajda, P.; Chang, S.-F. Brain state decoding for rapid image retrieval. In *Proceedings* of the 17th ACM International Conference on Multimedia; ACM: New York, NY, USA, 2009; pp. 945–954.
- Moon, J.; Kwon, Y.; Kang, K.; Bae, C.; Yoon, W.C. Recognition of Meaningful Human Actions for Video Annotation Using EEG Based User Responses. *Int. Conf. Multimed. Model.* 2015, 8936, 447–457. [CrossRef]
- Barry, R.J.; Clarke, A.R.; Johnstone, S.J.; Magee, C.A.; Rushby, J.A. EEG differences between eyes-closed and eyes-open resting conditions. *Clin. Neurophysiol.* 2007, 118, 2765–2773. [CrossRef]
- Tenke, C.E.; Kayser, J.; Alvarenga, J.E.; Abraham, K.S.; Warner, V.; Talati, A.; Weissman, M.M.; Bruder, G.E. Temporal stability of posterior EEG alpha over twelve years. *Clin. Neurophysiol.* 2018, 129, 1410–1417. [CrossRef] [PubMed]
- Huang, Z.; Weng, C.; Li, K.; Cheng, Y.-C.; Lee, C.-H. Deep learning vector quantization for acoustic information retrieval. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 1350–1354. [CrossRef]