



HAL
open science

Hyperparameter Optimization of Deep Learning Models for EEG-Based Vigilance Detection

Souhir Khessiba, Ahmed Ghazi Blaiech, Antoine Manzanera, Khaled Ben Khalifa, Asma Ben Abdallah, Mohamed Hédi Bedoui

► **To cite this version:**

Souhir Khessiba, Ahmed Ghazi Blaiech, Antoine Manzanera, Khaled Ben Khalifa, Asma Ben Abdallah, et al.. Hyperparameter Optimization of Deep Learning Models for EEG-Based Vigilance Detection. *Advances in Computational Collective Intelligence*, 1653, Springer International Publishing, pp.200-210, 2022, *Communications in Computer and Information Science*, 10.1007/978-3-031-16210-7_16 . hal-03835830

HAL Id: hal-03835830

<https://hal.science/hal-03835830v1>

Submitted on 1 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Hyperparameter Optimization of Deep Learning Models for EEG-Based Vigilance Detection

Souhir Khessiba^{1,2}[0000-0001-7351-7068], Ahmed Ghazi
Blaiech^{1,3}[0000-0002-9705-4042], Antoine Manzanera⁵[0000-0001-5718-411X],
Khaled Ben Khalifa^{1,3}[0000-0001-6758-1944], Asma Ben Abdallah^{1,4}, and
Mohamed Hédi Bedoui¹[0000-0003-4846-1722]

- ¹ Laboratoire de Technologie et Imagerie Médicale, Faculté de Médecine de Monastir,
Université de Monastir, 5019 Monastir, Tunisie
- ² Institut Supérieur d'Informatique et des Technologies de Communication de
Hammam Sousse, Université de Sousse, 4011, Sousse, Tunisie
- ³ Institut Supérieur des Sciences Appliquées et de Technologie de Sousse, Université
de Sousse, 4003
- ⁴ Institut supérieur d'informatique et de Mathématiques, Université de Monastir,
5019 Monastir, Tunisie
- ⁵ U2IS, ENSTA Paris, Institut Polytechnique de Paris, Palaiseau, France, 91120

Abstract. ElectroEncephaloGraphy (EEG) signals have a nonlinear and complex nature and require the design of sophisticated methods for their analysis. Thus, Deep Learning (DL) models, which have enabled the automatic extraction of complex data features at high levels of abstraction, play a growing role in the field of medical science to help diagnose various diseases, and have been successfully used to predict the vigilance states of individuals. However, the performance of these models is highly sensitive to the choice of the hyper-parameters that define the structure of the network and the learning process. When targeting an application, tuning the hyper-parameters of deep neural networks is a tedious and time-consuming process. This explains the necessity of automating the calibration of these hyper-parameters. In this paper, we perform hyper-parameters optimization using two popular methods: Tree Parzen Estimator (TPE) and Bayesian optimisation (BO) to predict vigilance states of individuals based on their EEG signal. The performance of the methods is evaluated on the vigilance states classification. Compared with empirical optimization, the accuracy is improved from 0.84 to 0.93 with TPE and from 0.84 to 0.97 with Bayesian optimization using the 1D-UNet-LSTM deep learning model. Obtained results show that the combination of the 1D-UNet encoder and LSTM offers an excellent compromise between the performance and network size (thus training duration), which allows a more efficient hyper-parameter optimization.

Keywords: Deep Learning models · Hyperparameter optimization · EEG · Vigilance.

1 Introduction

ElectroEncephaloGraphy (EEG) is the main modality for studying the electrical activity of the brain. However, the classification of these states from this signal requires sophisticated approaches in order to achieve the best performance. Deep Learning (DL) approaches have shown a good performance in learning the high-level features of signals [7] [18], particularly for EEG. They are characterized by their large number of hidden layers that provide the most effective solutions thanks to massive calculations.

One of the most powerful models in DL approaches is the Convolution Neural Network (CNN). Thus, many studies have suggested CNN models for analyzing the EEG signal. In [2], the authors utilized the concept of DL on EEG signals to predict the driver’s cognitive workload. A CNN model was used to extract features and accurately classify the cognitive workload. The experimental results showed that the proposed system could provide an accurate classification of high and low cognitive workload sessions. In [12], three types of deep covariance learning models were suggested to predict drivers’ drowsy and alert states using EEG signals: the CNN, the Symmetric Positive Definite Network (SPDNet), and the Deep Neural Network (DNN). The experimental results indicated that all the three models of deep covariance-learning reported a very good classification performance compared with shallow learning methods. In [14], the authors proposed two DL models to predict individuals’ vigilance states based on the study of one derivation of EEG signals: a 1D-UNet model and 1D-UNet-Long Short-Term Memory (1D-UNet-LSTM). Experimental results showed that the suggested models can stabilize the training process and well recognize the subject vigilance. Specifically, the per-class average of precision and recall could be respectively up to 86% with 1D-UNet and 85% with 1D-UNet-LSTM. All these studies have used several DL approaches to analyze EEG signals [12] [2] [14], but the choice of the architecture has been done empirically by the human expert through a slow trial and error process, guided mainly by intuition.

The success of the CNNs is highly dependent on the selection of the hyper-parameters. Determining which hyper-parameters to tune and defining value domains for those hyper-parameters, and then selecting the best set of values require meticulous design and experiment processes which can only be conducted by the participation of an expert from the domain. The need for the automatic design of CNNs is especially important for complex CNN architectures where the parameter space is so large that trying all possible combinations is computationally infeasible. Much research has been done in the field of Hyperparameter Optimization (HPO), such as grid search, random search, Bayesian optimization, and gradient-based optimization [9] [4]. Grid search and manual search are the most widely used strategies for HPO [9] [11]. These approaches make reproducibility harder and are impractical when there are a large number of hyper-parameters. Thus, many authors have focused on further automating the calibration of hyper-parameters. Particle Swarm Optimization (PSO) is among the metaheuristics that has been successfully applied for the optimization of CNN hyper-parameters. In [8], a parallel version of PSO algorithm was proposed for

the hyper-parameter optimization of DL models to overcome two problems: (i) the search space which is usually high dimensional, and (ii) the high runtime. The experiments have revealed that the PSO would largely take advantage of the rapidity offered by computational parallelization. Another PSO-based approach used to configure the CNN architecture was introduced in [15]. The parameters to be tuned are kernel size, padding, number of feature maps, and pooling patterns. By using the PSO adaptation, authors achieved better results than those found by AlexNet and got a high accuracy. In [10] an OLPSO (orthogonal Learning Particle Swarm Optimization) approach was presented in which hyperparameters values were optimized for both VGG16 and VGG19 networks for plant disease diagnosis. Batch size and dropout rate are employed as hyperparameters. Through practical experiments, the authors proved that their approach achieves higher performance and accuracy compared to other methods tested for the same data. The authors in [15] investigated lung nodule classification by proposing a multi-level CNN whose hyperparameter configuration was optimized by using a proposed Gaussian process with stationary kernels. The experiments demonstrated that the algorithm outperformed manual tuning. TPE algorithm in [17], has been proposed through Hyperas tool in order to optimize CNN hyperparameters to classify pulmonary nodules at an early stage. It was shown that the smallest, more basic CNN architecture, just one convolutional layer following of one max-pooling layer obtained the best results. The hyperas tool with the TPE algorithm allowed to explore all the hyperparameters in the experiments and it was important to achieve excellent results. Therefore, in this paper, we describe the use of two popular methods: TPE and BO for automatically designing and training DL models to predict individuals' vigilance states from an EEG signal. Those algorithms are applied on the 1D-UNet and 1D-UNet-LSTM models proposed in [14] to improve the classification performance.

This paper is structured as follows: Section 2 presents the materials and methods and introduces the DL models successfully implemented for vigilance state classification. It also defines TPE and BO optimization algorithm. Section 3 presents the data and the experimentation setup. Moreover, this section describes the results of the optimized suggested model and elaborates the discussion based on the obtained results. The last section concludes the paper and gives some future perspectives.

2 Materials and Methods

One of the most important strategies used to estimate vigilance consists in using physiological measures to give more precise data about the state of an individual. The sequential steps of the development of the automated vigilance state detection system are: EEG data collection, pre-processing and classification by DL, using hyperparameter optimization (see Fig. 1).

2.1 EEG signal acquisition and pre-processing

In this paper EEG signal is used to predict the vigilance states. This nonlinear and non-stationary signal characterizes the brain activity through a weakly in-

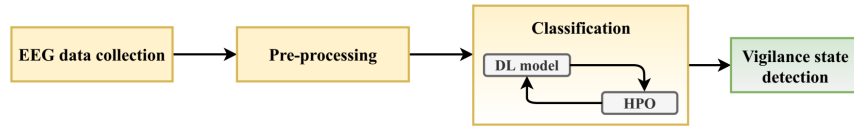


Fig. 1. EEG signal processing steps with HPO of DL models for vigilance state classification

vasive acquisition process, with electrodes placed along the scalp. To prepare the dataset, we use the same two subjects (S1, S2) as those collected in the experimentation of the previous work of our team [14] [5]. The EEG data are directly recorded from 28 active electrodes from the scalp at the Department of Functional Explorations of the Nervous System at Sahloul University Hospital, Tunisia. This signal is recorded during three 24h periods with a 15 days interval, and it involves two healthy male subjects aged between 18 and 23. For each subject, the signal is recorded for two states: vigilance state (VS) and drowsiness state (DS). The EEG recordings are done, reviewed and approved by an expert, in order to label the different levels of alertness. In this work, we focus on analyzing a single EEG signal from the right parieto-occipital (Pz-Oz) electrode used to characterize analyzed vigilance states. This choice is justified by the fact that experts agree that this signal is the most appropriate to reflect a state of vigilance and to enable the system portability [14] [5] [6]. In the first step of pre-processing, we split the signal into time periods of four seconds (recommended by an expert), in order to reduce the computation complexity. Then, we filter this signal to eliminate artifacts using a high-pass filter to remove low frequencies less than 0.1 Hz, and a low-pass filter to filter out frequencies above 21Hz, in order to focus on frequencies most related to the state of alertness. Experts agree that this range is one of the most relevant ranges for vigilance. The next step of pre-processing is the spectral analysis of the signal which was proposed in [14] [5]:

(i) The 512-point Fast Fourier Transform (FFT) is used to map the acquired time-series EEG data $f(t)$ to the frequency domain $F(u)$.

(ii) The frequency range [0.1 Hz, 21 Hz], which is specific to the range of physiological waves, is split into k elementary frequency bands to characterize this electrical activity.

(iii) In each band $[u_i, u_{i+1}]$, the Spectral Band Power (PBS), which corresponds to the sum of the spectral amplitudes belonging to the frequency band, is calculated:

$$PBS_i = \sum_{u \in [u_i, u_{i+1}]} \|F(u)\|; \quad \begin{cases} u_i = 0.1 + (i - 1) * \Delta u; i \in [1..k] \\ \Delta u = \frac{(21-0.1)}{k} \end{cases}$$

(iv) The Percentage of the Relative Spectral Power (PRSP) of each band is computed, which is equal to the PBS divided by the total spectral power:

$$PRSP_i = \frac{PBS_i}{TSP} \times 100, \text{ with } TSP = \sum_{u \in [0.1, 21]} \|F(u)\|$$

where Δu is the length of the frequency band (Hz), k is the number of bands, and TSP represents the total spectral power. Thereby, the PRSP will be the input to the classification tool for vigilance state detection, for each four second time sample.

2.2 DL models hyperparameters

DL models are widely applied to various areas like computer vision, classification and segmentation, since they have had great success solving many types of highly complex problems. Among the most powerful models in DL approaches are the CNN, in particular the 1D-CNN which has been well adopted in the literature for processing EEG signals [16]. Its architecture is usually composed by a series of 1D convolutional, pooling, normalization and fully connected layers. In this paper, we use the 1D-UNet-LSTM DL model recently proposed in [14] and successfully implemented for vigilance state classification.

1D-UNet-LSTM: The model presented in (Fig. 2) is a combination between 1D-UNet and LSTM.

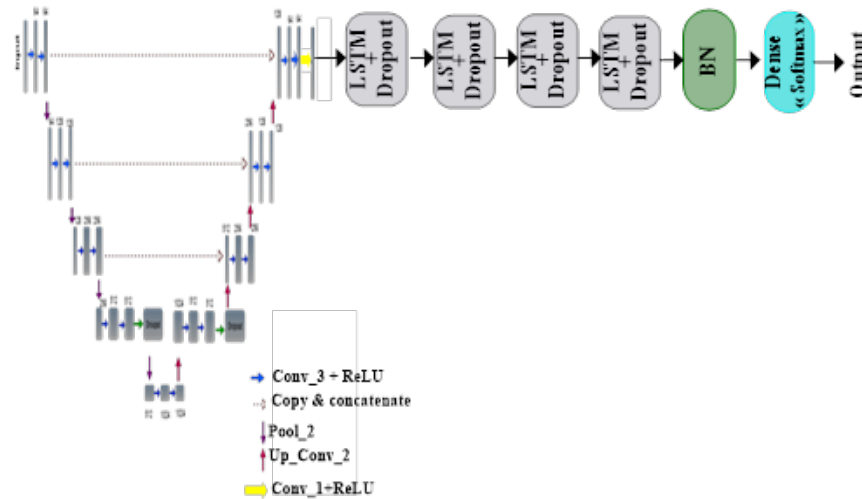


Fig. 2. 1D-UNet-LSTM architecture

The 1D-UNet-LSTM architecture takes the output of 1D-UNet (last layer) to feed in as the input of the LSTM network. This latter is made up of five hidden cells, where each cell is followed by a dropout layer to prevent overfitting. At the

end, the 1D-UNet-LSTM architecture integrates a batch normalization layer, a fully connected (dense) layer and Softmax layers to accomplish the classification task.

Hyperparameters: The DL neural network have many hyperparameters, including those that specify the structure of the network itself and those that determine how the network is trained. As the training of these networks is slow, it is difficult to adjust the hyperparameters. When training a network, the result of the model will depend not only on the chosen structure but also on the training method, which itself has several hyperparameters such as the learning rate, the loss function, the mini-batch size, and the number of training iterations. Furthermore, the structure of the neural network itself involves numerous hyperparameters in its design, including the size of each layer, the number of hidden layers, the number of convolution layers, the kernel size, the filter size, the activation function, the weight initialization, etc. Table 1 summarizes the hyperparameters responsible for defining the structure of the network and those related to the optimization and training process. Tuning the hyperparameters of DL neural network is a critical and time-consuming process that has been mainly done relying on the knowledge of the experts. This explains the necessity of automating the calibration of these hyperparameters.

Table 1. Hyperparameters defining architectures (top) and training process (bottom) of the neural network.

Hyperparameters	Types	Scope
Number of convolution layers	Integer	0,1,...,25
Number of LSTM layers	Integer	0,1,...,25
Number of dense layers	Integer	0,1,...,25
LSTM units	Integer	32,...,512
Optimizer	Categorical/Integer	Adam, Rmsprop, Adadelata
Filter size	Integer	64,128,...,1024
Kernel Size	Integer	0,...,10
Batch size	Integer	10, 32, 64,128
Learning rate	Float	0;1
Dropout rate	Float	0;1
Activation function	Categorical	Relu, Sigmoid, Tanh

2.3 HPO Algorithms

Deep model design requires strong knowledge of algorithms and appropriate hyperparameter optimization techniques. Several methods have been proposed for HPO such as grid search [9], random search, simulated annealing [?], BO [19] and TPE [3]. The TPE and BO success in expensive optimization problems indicates that they may outperform existing methods.

TPE algorithm is a Sequential Model-Based Optimization (SMBO) approach. SMBO methods sequentially construct models to approximate the performance

of hyperparameters based on historical measurements, and then choose new hyperparameters to be tested based on this model. Consequently, the TPE is an iterative process that uses the history of evaluated hyperparameters to create a probabilistic model, which is used to suggest the next set of hyperparameters to evaluate. Let assume a set of observations $\{(x^{(1)}, y^{(1)}), \dots, (x^{(k)}, y^{(k)})\}$, To apply the TPE, the observation results are divided into good and poor results by a pre-defined percentile y^* . The TPE defines $p(x | y)$ using the following two probability density functions given by the equation:

$$P(x | y) = \begin{cases} l(x) & \text{if } y < y^* \\ g(x) & \text{if } y \geq y^* \end{cases}$$

Where $l(x)$ is the probability density function formed using the observed variables $\{x^{(i)}\}$ such that $y^* > y^{(i)}$ ($= f(x^{(i)})$), and $g(x)$ is the probability density function using the remaining observations. Value y^* is selected to be a quantile γ of the observed y values satisfying $p(y^* > y) = \gamma$. After that, the expected improvement in the acquisition function is reflected by the ratio between the two density functions, which is used to determine the new configurations for evaluation.

BO algorithm tries to minimize a scalar objective function $f(x)$ for x . Depending on whether the function is deterministic or stochastic, the output will be different for the same input x . The minimization process comprises three main components: a Gaussian process model for the objective function $f(x)$, a Bayesian update process that modifies the Gaussian model after each new evaluation of the objective function, and an acquisition function $a(x)$. This acquisition function is maximized in order to identify the next evaluation point. The role of this function is to measure the expected improvement in the objective function while discarding values that would increase it. Hence, the expected improvement (EI) is calculated as:

$$EI(x, Q) = E_Q [\max(0, \mu_Q(x_{\text{best}}) - f(x))]$$

where Q is the posterior distribution function, x_{best} is the location of the lowest posterior mean and $\mu_Q(x_{\text{best}})$ is the lowest value of the posterior mean.

Compared to a grid search or manual tuning, BO allows us to jointly tune more parameters with fewer experiments and find better values[13].

3 Experiments and results

The implementation has been done to show the effectiveness of the HPO algorithm used to improve the performance of vigilance state classification.

3.1 Experiment setting

We evaluate the hyperparameter optimization algorithms on the 1D-UNet-LSTM architecture. This architecture is developed using Keras whose libraries are written in Python. The experiments are achieved with an experimental implementation on a Pop Gaming laptop PC with an Intel 9th-generation Core i5-9300H

processor, a NVIDIA GeForce GTX 1650 Graphics card and 8 GB Memory. To tackle HPO problems, we use Optuna framework [1], which provides many HPO algorithms including the TPE and Sherpa [20] framework which provides BO algorithm.

3.2 Results and discussion

This section describes the results obtained. We focus on two subjects [14] [5] with the same size of observations in order to detect the vigilance states. The within-subject vigilance state classification is applied to evaluate the performance by different models, where each subject is taken separately and divided into 80% and 20% of observations for training and testing, respectively. Table 2 presents the hyper-parameter values obtained by the implemented DL models for the two subjects. This table shows that Adam function is more often selected as an optimizer, which justifies the effectiveness of this function. Furthermore, the ReLU activation is selected for all implementations. We note that the hyper-parameter values change between the models for the same subject. This proves that the hyperparameters are specific to the utilized architecture. Furthermore, the hyperparameter values vary between the subjects within the same DL model. This proves also that the hyperparameters depend on the input data, even if we work in the same context.

Table 2. Best hyperparameters configurations using TPE and BO algorithms

	1D UNET-LSTM			
	TPE		BO	
	S1	S2	S1	S2
Number of convolution layers	10	10	9	13
Number of LSTM layers	4	5	5	7
LSTM Units	100	64	150	125
Optimiser	Adam	Adam	Adam	Adam
Filter size	64	32	128	64
Kernel size	1	1	1	1
Batch size	10	10	64	10
Learning rate	0.002	0.003	0.002	0.001
Dropout rate	0.4	0.3	0.3	0.5
Activation function	Relu	Relu	Relu	Relu

Table 3 exposes the accuracy results obtained using HPO algorithms and compared with the results before the optimization process for the 1D-UNet-LSTM model. We note that the classification performance in terms of accuracy is good using TPE and BO algorithms. Accuracy for subject S1 using TPE can be up to 0.93 with 1D-UNet-LSTM. and with BO, Accuracy can be up to 0.97. The accuracy gain with TPE can reach 12.5% and with BO can reach 15.51% for subject S2 using 1D-UNet-LSTM compared to an implementation without optimization.

Table 3. Subject vigilance state classification Accuracy

	S1					S2				
	Without HPO [2]	HPO-TPE	HPO-BO	Gain-TPE (%)	Gain-BO(%)	Without HPO [2]	HPO-TPE	HPO-BO	Gain-TPE (%)	Gain-BO(%)
1D Unet-LSTM	0.840	0.932	0.973	9.8	13.6	0.735	0.840	0.870	12.5	15.51

Table 4 describes the classification performance in terms of recall, precision and F1-score using UNet-LSTM architecture for subject S1, which has the best classification accuracy, as depicted in Table 2 (the per-model average is 0.956 using BO). This table shows that the precision can achieve 0.92 using 1D-UNet-LSTM with HPO. The Precision gain is 13.41% using 1D-UNet-LSTM with HPO (BO algorithm) compared to the same model without optimization.

Table 4. Performance measures of proposed models for subject 1

	Recall					Précision					F1-Score				
	Without HPO [2]	HPO-TPE	HPO-BO	Gain-TPE (%)	Gain-BO(%)	Without HPO [2]	HPO-TPE	HPO-BO	Gain-TPE (%)	Gain-BO(%)	Without HPO [2]	HPO-TPE	HPO-BO	Gain-TPE (%)	Gain-BO(%)
1D Unet-LSTM	0.85	0.89	0.902	4.4	5.7	0.80	0.90	0.924	11.1	13.41	0.81	0.894	0.912	9.3	11.18

Given Table 3 and Table 4, we note that including an optimization phase of hyperparameters allows to significantly improve the classification performance for all subjects and for all implemented DL model. Indeed, these results show that the iterative process of BO and TPE are suitable for our application.

4 Conclusion and Perspectives

In this paper, we have introduced and explored the potential of HPO algorithms in order to give the best configurations of hyperparameters and to improve the performance of vigilance state classification EEG signals. The HPO TPE and BO have been applied to the 1D-UNet-LSTM model, and the optimal hyperparameter configuration has been generated. The experimental results in the study have revealed that the performance of vigilance state classification has been improved using the HPO BO method and the accuracy gain can reach 15.51% for subject S2 using 1D-UNet-LSTM compared to an implementation without an optimization process. In the future, we will add more subjects for further validation of the DL architecture with hyperparameter optimization. In addition, we will evaluate more HPO algorithms in order to improve the system performance.

References

1. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery data mining. pp. 2623–2631 (2019)
2. Almogbel, M.A., Dang, A.H., Kameyama, W.: Eeg-signals based cognitive workload detection of vehicle driver using deep learning. In: 2018 20th International Conference on Advanced Communication Technology (ICACT). pp. 256–259. IEEE (2018)
3. Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B.: Algorithms for hyper-parameter optimization. *Advances in neural information processing systems* **24** (2011)
4. Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization. *Journal of machine learning research* **13**(2) (2012)
5. Blaiech, A.G., Ben Khalifa, K., Boubaker, M., Bedoui, M.H.: Lvq neural network optimized implementation on fpga devices with multiple-wordlength operations for real-time systems. *Neural Computing and Applications* **29**(2), 509–528 (2018)
6. Blaiech, A.G., Khalifa, K.B., Boubaker, M., Bedoui, M.H.: Multi-width fixed-point coding based on reprogrammable hardware implementation of a multi-layer perceptron neural network for alertness classification. In: 2010 10th International Conference on Intelligent Systems Design and Applications. pp. 610–614. IEEE (2010)
7. Boudegga, H., Elloumi, Y., Akil, M., Bedoui, M.H., Kachouri, R., Abdallah, A.B.: Fast and efficient retinal blood vessel segmentation method based on deep learning network. *Computerized Medical Imaging and Graphics* **90**, 101902 (2021)
8. Brito, R., Fong, S., Zhuang, Y., Wu, Y.: Generating neural networks with optimal features through particle swarm optimization. In: Proceedings of the International Conference on Big Data and Internet of Thing. pp. 96–101 (2017)
9. Claesen, M., De Moor, B.: Hyperparameter search in machine learning. *arXiv preprint arXiv:1502.02127* (2015)
10. Darwish, A., Ezzat, D., Hassanien, A.E.: An optimized model based on convolutional neural networks and orthogonal learning particle swarm optimization algorithm for plant diseases diagnosis. *Swarm and evolutionary computation* **52**, 100616 (2020)
11. Firdaus, F.F., Nugroho, H.A., Soesanti, I.: Deep neural network with hyperparameter tuning for detection of heart disease. In: 2021 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob). pp. 59–65. IEEE (2021)
12. Hajinoroozi, M., Zhang, J.M., Huang, Y.: Driver’s fatigue prediction by deep covariance learning from eeg. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC). pp. 240–245. IEEE (2017)
13. Injadat, M., Salo, F., Nassif, A.B., Essex, A., Shami, A.: Bayesian optimization with machine learning algorithms towards anomaly detection. In: 2018 IEEE global communications conference (GLOBECOM). pp. 1–6. IEEE (2018)
14. Khessiba, S., Blaiech, A.G., Ben Khalifa, K., Ben Abdallah, A., Bedoui, M.H.: Innovative deep learning models for eeg-based vigilance detection. *Neural Computing and Applications* **33**(12), 6921–6937 (2021)
15. Khoong, W.H.: A heuristic for efficient reduction in hidden layer combinations for feedforward neural networks. In: Science and Information Conference. pp. 208–218. Springer (2020)
16. Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., Inman, D.J.: 1d convolutional neural networks and applications: A survey. *Mechanical systems and signal processing* **151**, 107398 (2021)

17. Konar, J., Khandelwal, P., Tripathi, R.: Comparison of various learning rate scheduling techniques on convolutional neural network. In: 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). pp. 1–5. IEEE (2020)
18. LeCun, Y.: Yoshua bengio, and geoffrey hinton. Deep learning. *nature* **521**(7553), 436–444 (2015)
19. Snoek, J., Larochelle, H., Adams, R.P.: Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems* **25** (2012)
20. Wu, J., Chen, X.Y., Zhang, H., Xiong, L.D., Lei, H., Deng, S.H.: Hyperparameter optimization for machine learning models based on bayesian optimization. *Journal of Electronic Science and Technology* **17**(1), 26–40 (2019)