IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 18 July 2022, accepted 21 September 2022, date of publication 5 October 2022, date of current version 19 October 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3212403

RESEARCH ARTICLE

A Dopamine Based Adaptive Emotional Neural Network

MOHAMMAD AMIN ZARE[1], REZA BOOSTANI[1], MOKHTAR MOHAMMADI[2], AND SAMANEH KOUCHAKI[3]

[1]Department of Computer Sciences and Engineering, Shiraz University, Shiraz 71348-14336, Iran
[2]Department of Information Technology, College of Engineering and Computer Science, Lebanese French University, Kurdistan Region, Erbil 44001, Iraq
[3]Centre for Vision, Speech, and Signal Processing, Department of Electrical and Electronic Engineering, University of Surrey, Guildford, Surrey GU2 7XH, U.K.

Corresponding author: Samaneh Kouchaki (samaneh.kouchaki@surrey.ac.uk)

ABSTRACT Due to the inevitable role of emotions in human learning and decision-making, different types of emotions in the form of emotional weights/neurons have also been considered in shallow neural networks. Emotional neural networks suffer from a low convergence rate as well as batch learning instability mainly because of the improper tuning of learning coefficients. To overcome these drawbacks, we introduced two solutions: (i) a heuristic upgrading method, inspiring by the behavior of dopamine secretion in the human brain, to adaptively regulate the learning rate based on positive and negative emotional states at each epoch and (ii) a stochastic learning technique to stabilize the learning process. The proposed dopamine based adaptive emotional neural network statistically outperforms state-of-the-art methods like emotional neural network, prototype-incorporated emotional neural network, multi-layer perceptron, and deep convolutional neural networks such as LeNet, AlexNet, DenseNet, MobileNet and EfficientNet in terms of different measures such as accuracy and convergence rate on several high dimensional and big datasets.

INDEX TERMS Adaptive learning rate, dopamine behavior, emotional neural network, shallow neural network.

I. INTRODUCTION

Emotions highly affect the learning process, cognition, and decision-making in both humans and animals [1], [2], [3], [4], [5], [6]. They are intelligent and complex adaptive systems and play critical roles in responding to humans' unpredictable and complex behaviors [1], [6], [7]. Therefore, incorporating the mechanism of emotion into artificial intelligence (AI) algorithms has gained attention in order to enhance their performance.

Researchers have claimed that most decisions and cognitions in humans are based on a mixture of intelligence and emotion [1], [4], [6], [8], [9]. Therefore, incorporating emotions such as nervousness, agony, despondency, anxiety or confidence in intelligent machines or artificial neural networks (ANN) can improve their performance in complex decision-making situations [10], [11].

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

ANN mimic the human brain's neural structure and simulate the neurons' function to learn patterns, predict a target, or make decisions [12]. Nonetheless, the actual mechanism of the brain network is more complex than ANNs because the functionality of all neurons is not the same [6]. Moreover, the role of different hormones activities like dopamine in learning the synaptic weights is mostly ignored in ANNs [13]. Several attempts have been made to equip ANNs with human emotions and hormones activities [14], [15], [16]. Researchers considered anxiety and confidence as genuine human emotions mainly because these emotional states positively/negatively affect learning and decision-making in humans [11], [15], [17]. As an example, an emotional neural network (EmNN) was proposed in [15]. This network adopts two types of emotional states (anxiety and confidence), emotional neurons and backpropagation (BP) learning algorithm called EmBP to update the weights in batch learning. The results of EmNN have illustrated more improvement in learning and decision-making than conventional multi-layer perceptron (MLP) in complex situations.

Additionally, the root of using anxiety and confidence in EmNN emerges from the fact that when children learn a new task, they initially show a high-level of anxiety and low-level of confidence. However, after practicing and getting positive feedback during learning, anxiety decreases while the level of confidence increases [2], [6], [10], [11], [15].

Another emotional neural network called prototype-incorporated emotional neural network (PI-EmNN) was proposed in [18]. They incorporated prototype and adaptive learning theories into EmNN to improve the overall learning and decision-making process. Furthermore, this model employs only one hidden layer and no convolution operations for feature learning like EmNN.

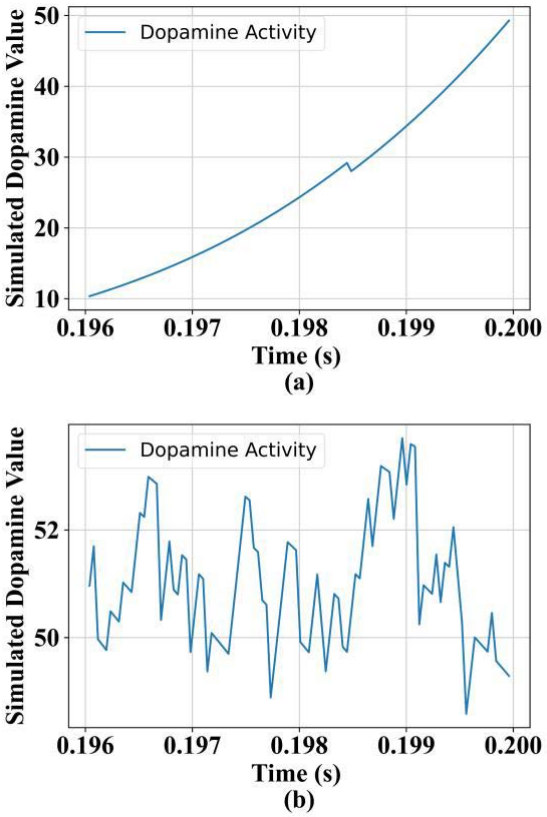


FIGURE 1. The simulated dopamine activities (DA) in 0.002 seconds (a) before or after learning (b) during learning based on the real dopamine changes in humans [19].

Their model has shown competitive performance compared to deep convolutional neural networks (CNNs) with many hidden layers and convolution operations. While the usage of emotions in EmBP and PI-EmNN expedites and improves learning and decision-making processes compared to conventional ANNs, they are too slow to handle large datasets. Additionally, because they employ batch learning, which is sensitive to even modest changes in learning rate, their outcomes are associated with the learning rate [25], [26].

On the other hand, various neurotransmitters or hormones have vital roles in learning, cognition, and decision-making

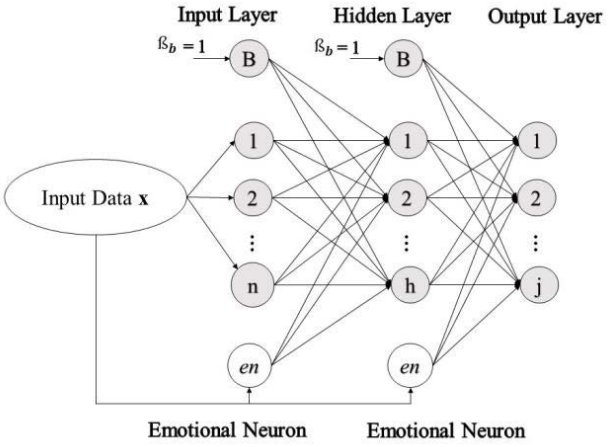


FIGURE 2. The three-layer structure of the proposed Adaptive Emotional Neural Network (AEmNN) has emotional neurons *en*, input, hidden and output layer with bias $ß_b$ which is initialized to one.

in the human and animal brains, and it is widely accepted that the dopamine neurotransmitter has a special role in learning [21], [27], [28]. Additionally, the presence of dopamine is highly required for learning, decision-making, and behavioral control, and its absence can cause a variety of diseases such as Parkinson's, Schizophrenia, addiction, Alzheimer's and depression [20]. Hence, modeling the behavior of dopamine and formulating its variations into synaptic weights' learning can enhance the performance of ANNs [24], [29], [30], [31], [32], [33], [34], [35].

Moreover, cutting-edge [24], [29], [30], [31], [32], [33], [34], [35] studies have demonstrated that dopamine fluctuations in humans and animals can encode reward prediction error (RPE) in learning. RPE indicates the difference between actual and expected outcomes and originates from earlier learning theories of machine learning [36]. The recent studies [23], [27], [37] conducted experiments on the dopamine fluctuation behavior of animals and humans. They demonstrated that dopamine uses reward and punishment factors to determine its level which causes fluctuations during learning a new task. These factors are obtained from the performance (difference between actual and expected outcomes) of humans and animals during learning. The reward rate can dynamically change the level of dopamine when the actual and expected outcomes are the same. Moreover, the punishment rate can dynamically decrease its level when the actual and expected outcomes are not the same. These activities create numerous fluctuations during learning a new task that leads to learning in humans and animals. On the contrary, when learning a new task does not occur, dopamine levels are stationary or continuously increasing or decreasing as can be seen in recent studies or the simulated dopamine fluctuations in Fig. 1 (a).

Other studies [20], [21], [38], [39], [40], [41] worked on the interactions between positive/negative human emotions and human dopamine levels. They used men and women in

their experiments. They found out that human performance (difference between actual and expected outcomes) can create positive and negative emotions, which can dynamically change dopamine levels. When their performance was good (actual and expected outcomes are the same), the positive feeling controlled the dopamine level, and when their performance was not good (actual and expected outcomes are not the same), the negative feeling controlled the dopamine level.

TABLE 1. Training algorithm and device required to execute each model.

Model	Learning Algorithm	Run with	Learning Mechanism
AEmNN	EmBP	CPU/GPU	SGD-adaptive
MLP	BP	CPU	BGD
EmNN	EmBP	CPU	BGD
PIEmNN	EmBP	CPU	BGD
LeNet-5	BP	GPU	Mini-Batch SGD
AlexNet	BP	GPU	Mini-Batch SGD
EfficientNet-B0	BP	GPU	Mini-Batch SGD
MobileNet-V2	BP	GPU	Mini-Batch SGD
DenseNet-121	BP	GPU	Mini-Batch SGD

To improve the existing emotional networks, we suggest a new adaptive emotional neural network (AEmNN). This network uses an adaptive emotional learning rate. The adaptive emotional learning rate mimics the human dopamine fluctuations behavior based on studies in [23], [27], [37], [20], [21], and [38]. The learning rate value is determined by the two emotional states of AEmNN. These states control the learning rate in each epoch and create fluctuations within a specified range based on AEmNN performance. This process is similar to the actual behavior of dopamine fluctuation in humans. It is noteworthy that this procedure adaptively regulates the learning rate coefficient at each epoch to speed up the convergence and help the network learn more efficiently and make improved decisions than others.

AEmNN directly affects the learning steps of artificial neurons and guide learning. In contrast, in the former approaches, emotional neurons just manipulate artificial neurons and the network's bias. The proposed AEmNN is applied to several facial recognition datasets and is compared to MLP [42], EmBP [15], [43], PI-EmNN [18], LeNet-5 [44], AlexNet [45], DenseNet-121 [46], MobileNet-V2 [47] and EfficientNet-B0 [48]. Besides, Table 1 shows their learning mechanisms. To assess the methods' performance, accuracy, precision, recall, f-measure (f1-score), r-squared, mean absolute error (MAE), convergence rate, and statistical test are determined, and the evaluation is performed using ten-fold cross-validation considering ten independent runs.

The rest of this paper is organized as follows: Section II introduces the proposed adaptive emotional neural network (AEmNN). Section III describes the characteristics of the employed datasets. Section IV illustrates the training algorithms and settings, comparative results and discusses the pros and cons of the achieved results by our scheme compared to state-of-the-art techniques (MLP, EmNN, PI-EmNN, LeNet-5, AlexNet, DenseNet-121, MobileNet-V2 and EfficientNet-B0). Finally, Section V concludes the paper and gives a horizon to the future of this work.

II. METHODS

Like EmNN, the proposed AEmNN contains emotional-weights, -neurons, and -parameters (anxiety and confidence), but uses a sigmoid activation function with cross-entropy loss. Fig. 2 illustrates the structure of our network. One of our method's main distinguishing properties is the use of stochastic learning instead of batch learning backpropagation to update its weights for each incoming sample and stabilize the learning process. We note that stochastic learning, equipped with adaptive emotional learning rates, can preserve both exploitation and exploration suitably and stably. Moreover, several research studies have empirically shown that stochastic learning cannot be destabilized by adaptive changing the learning rate [25], [26]. It is faster and more accurate than batch learning, especially in dealing with big data. The proposed method initialized the confidence and anxiety parameters similar to PI-EmNN. Then, AEmNN adaptively changes the learning rate based on dopamine fluctuations in the human brain, the linear combination of the feedback error and the emotional parameters in each epoch. Adaptively changing the learning rate is inspired especially by the study in [20] and [49]. This implies that the emotional parameters can significantly and directly impact and update the weights of artificial neurons.

The adaptive learning rate enables stochastic learning that is not sensitive to the learning rate's initialization to explore and exploit the search space [41]. The idea of changing the learning rate with two fixed step sizes (randomly initialized at first) was suggested by [41] and [42]. However, the conventional methods suffer from the absence of a flexible instance-based learning scheme to adjust the learning coefficients adaptively. To solve this problem, we have developed a heuristic method based on the process of dopamine changes (see Fig. 1 (b)) in the human/animal brain [24], [33], [40], [43], [44], [45], [46], to regulate the learning rate and adjust it based on emotional parameters in each epoch.

As shown in Fig. 2, prior knowledge (initial values of the emotional neurons) was elicited from input data. The emotional neuron *en* is a visual stimulus that evokes emotions (positive or negative) [15], [50] and initializes by taking a global average value over each input pattern **x** with n dimensions, as described in (1):

$$en = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

The output of each hidden neuron h denoted as y_h and the output (activation) of any hypothetical output neuron j denoted as y_j are computed as follows:

$$y_h = f\left(\sum_{i=1}^{n} w_{hi}\cdot x_i + w_{hb}\cdot ß_b + w_{he}\cdot en\right) \tag{2}$$

TABLE 2. Description of the employed datasets.

Dataset	Train Size	Test Size	#Classes	Image description
ORL	360	40	40	Different times, varying the lighting, facial expressions (opened / closed eyes, smiling / not smiling) and facial details (glasses / no glasses)
Yale	148	17	15	Different facial expressions or configurations: center-light, glasses, happy, left-light, no glasses, normal, right-light, sad, sleepy, surprised, and wink
Yale-B	2,189	243	38	Illumination conditions
MIT	4716	524	9	Vary the illumination, pose (up to about 30 degrees of rotation in-depth), and the background
MNIST	60,000	10,000	10	Handwritten digits from 0 to 9
Fashion-MNIST	60,000	10,000	10	Trouser, pullover, dress, coat, sandal, shirt, sneaker, bag, and ankle boot
CIFAR-10	50,000	10,000	10	Colored natural images with 32×32 pixels
CIFAR-100	50,000	10,000	100	Colored natural images with 32×32 pixels
SVHN	73,257	26,032	10	32×32 colored digit images
CINIC-10	180,000	90,000	10	An extension of CIFAR-10 via the addition of downsampled ImageNet 32×32 colored images.

TABLE 3. Fine-tuned parameters, settings, training parameters and run time of each network for each dataset achieved through ten-fold cross-validation on a train set. AEmNN was executed by GPU for datasets including CIFAR-10, CIFAR-100, SVHN and CINIC-10. * indicates GPU execution time.

Dataset	Model	Initial Learning Rate	Momentum	Minimum Train MSE/ CE Loss	Training Time (s)	Training Parameters	Net Structure	Final Anxiety	Final Confidence
ORL	AEmNN	0.01	0.22	0.0017	16	76 T	[1025,72,40]	0.003	0.52
	MLP	0.6	0.21	0.002	21	9 T	[100,60,40]	-	-
	EmNN	0.5	0.002	0.002	28	9.5 T	[101,62,40]	0.007	0.33
	PI-EmNN	0.008	0.002	0.0017	19	79 T	[1027,74,40]	0.006	0.44
Yale	AEmNN	0.01	0.021	0.003	4.9	32 T	[1025,32,15]	0.003	0.48
	MLP	0.2	0.0021	0.003	9.1	3.5 T	[100,30,15]	-	-
	EmNN	0.09	0.0021	0.003	8.4	3.7 T	[101,32,15]	0.009	0.27
	PI-EmNN	0.05	0.002	0.003	11	76 T	[1027,74,15]	0.009	0.36
Yale-B	AEmNN	0.01	0.22	0.012	14	64 T	[1025,62,38]	0.01	0.53
	MLP	0.05	0.31	0.002	40	8.4 T	[100,60,38]	-	-
	EmNN	0.6	0.3	0.002	37	8.5 T	[101,62,38]	0.005	0.31
	PI-EmNN	0.8	0.21	0.002	23	78 T	[1027,74,38]	0.004	0.35
MIT	AEmNN	0.01	.021	0.003	32	32 T	[1025,32,10]	0.003	0.33
	MLP	0.9	0.0021	0.003	96	3.3 T	[100,30,10]	-	-
	EmNN	0.85	0.0021	0.003	80	3.5 T	[101,32,10]	0.007	0.25
	PI-EmNN	0.39	0.002	0.003	47	75 T	[1027,74,10]	0.006	0.26
MNIST	AEmNN	0.01	0.024	0.1	343	156 T	[1025,152,10]	0.15	0.18
	MLP	0.9	0.2	0.02	501	55 T	[100,500,10]	-	-
	EmNN	0.69	0.31	0.02	763	56 T	[101,502,10]	0.1	0.19
	PI-EmNN	0.3	0.002	0.02	1010	211 T	[1027,204,10]	0.05	0.34
Fashion-MNIST	AEmNN	0.01	0.02	0.15	502	156 T	[1025,152,10]	0.17	0.25
	MLP	0.45	0.2	0.02	618	55 T	[100,500,10]	-	-
	EmNN	0.69	0.31	0.02	802	56 T	[101,502,10]	0.3	0.4
	PI-EmNN	0.73	0.002	0.02	959	211 T	[1027,204,10]	0.35	0.42
CIFAR-10	AEmNN	0.01	0.9	1.3	1800 *	0.8 M	[1025,802,10]	0.93	0.59
CIFAR-100	AEmNN	0.01	0.9	2.8	2000 *	0.9 M	[1025,802,100]	2.4	1.4
SVHN	AEmNN	0.01	0.9	0.89	1200 *	0.8 M	[1025,802,10]	0.52	1.15
CINIC-10	AEmNN	0.01	0.9	0.2	2580 *	0.8 M	[1025,802,10]	0.2	1.6

$$y_j = f\left(\sum_{h=1}^{l} w_{jh} \cdot y_h + w_{jb} \cdot ß_b + w_{je} \cdot en\right) \quad (3)$$

where x_i is the i^{th} input feature, w_{hi} is the weight between input neuron i and hidden neuron h, w_{jh} is the weight between hidden neuron h and output neuron j. w_{hb} is the weight between the bias neuron and hidden neuron h, w_{jb} is the bias weight for the j^{th} output neuron, $ß_b$ is the bias of the hidden and output layer, which is initialized to 1. w_{he} is the interconnection weight from emotional input neuron en to the hidden layer neuron h, w_{je} is the emotional weight from hidden emotional neuron en to the j^{th} output neuron, n is the number of features, and l *is* the total number of neurons in the hidden layer, f is the sigmoid activation function and calculated as follows:

$$f(x) = 1/((1 + \exp(-x)) \quad (4)$$

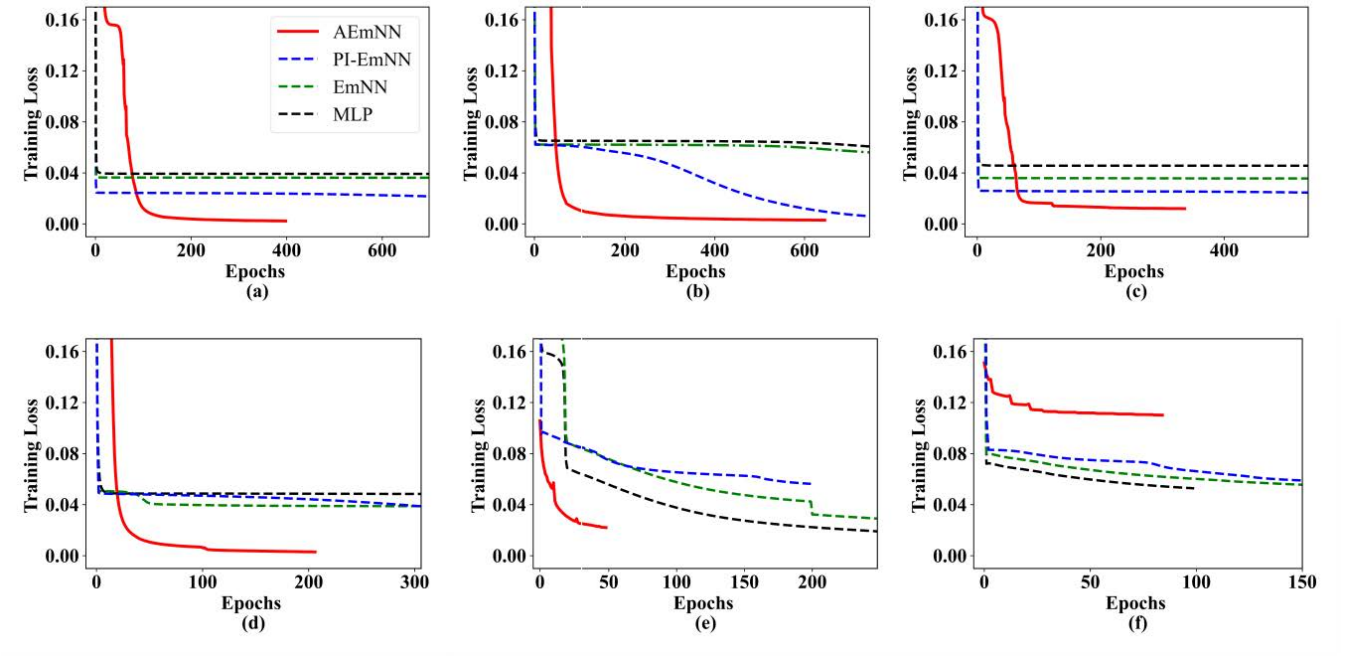


FIGURE 3. The learning curve of (a) ORL, (b) Yale, (c) Yale-B, (d) MIT, (e) MNIST, and (f) Fashion-MNIST fed to AEmNN, MLP, EmNN, and PI-EmNN.

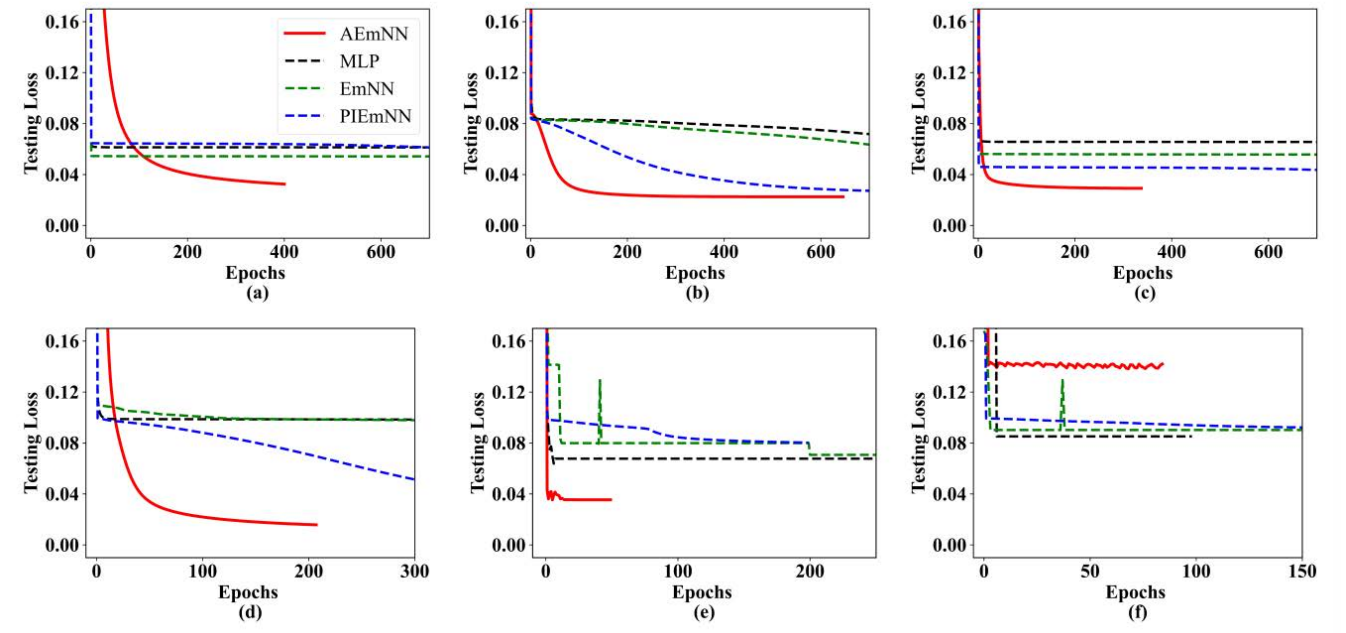


FIGURE 4. The testing loss curve of (a) ORL, (b) Yale, (c) Yale-B, (d) MIT, (e) MNIST, and (f) Fashion-MNIST fed to AEmNN, MLP, EmNN, and PI-EmNN.

The Network employs cross-entropy loss function e to calculate its loss as follows:

$$e = -\sum_{j=1}^{N_{classes}} t_j * \log(y_j) \tag{5}$$

where $N_{classes}$ is the number of output neurons, t_j and y_j are the target and output value at neuron j. The proposed Network uses root mean squared error (RMSE) at i^{th} epoch analogous to the state-of-the-art EmNN. RMSE helps the network to compute the anxiety and denoted as e_i:

$$rmse = \sqrt[2]{\sum_{m=1}^{p} (t_m - y_m)^2 \Big/ p} \tag{6}$$

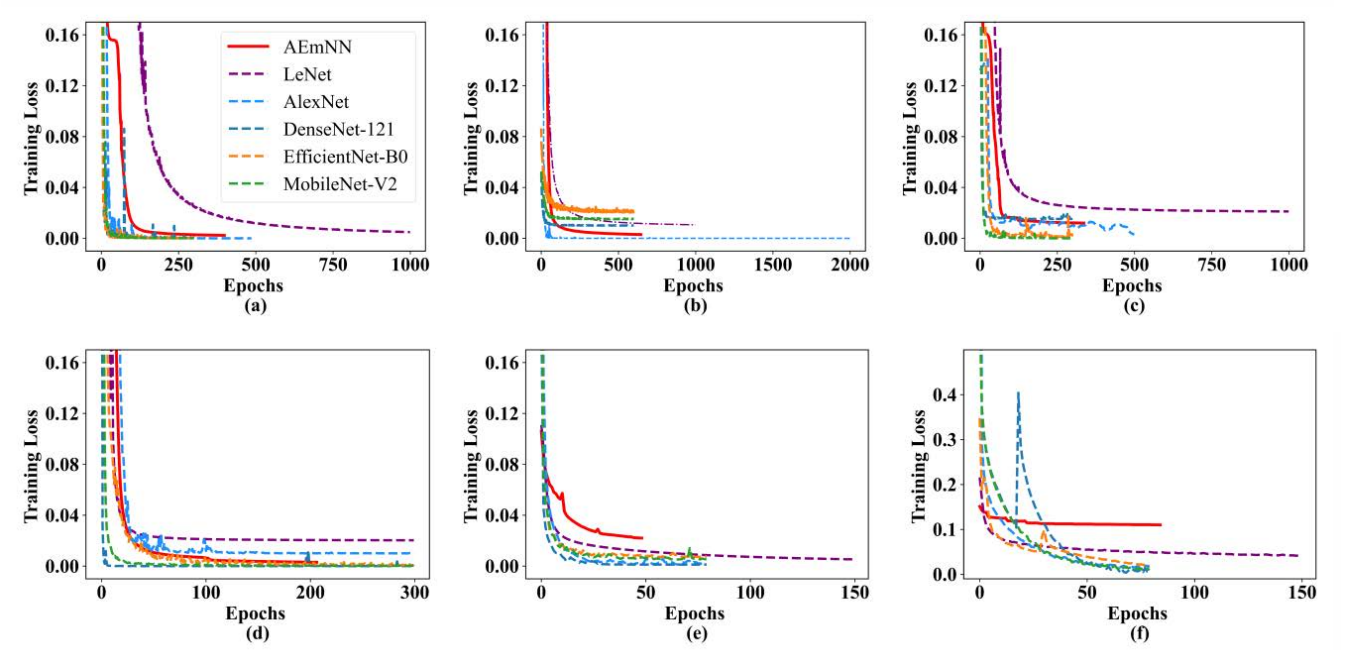


FIGURE 5. The learning curve of (a) ORL, (b)Yale, (c) Yale-B, (d) MIT, (e) MNIST, and (f) Fashion-MNIST fed to AEmNN, LeNet, AlexNet, DenseNet-121, EfficientNet-B0 and MobileNet-V2.

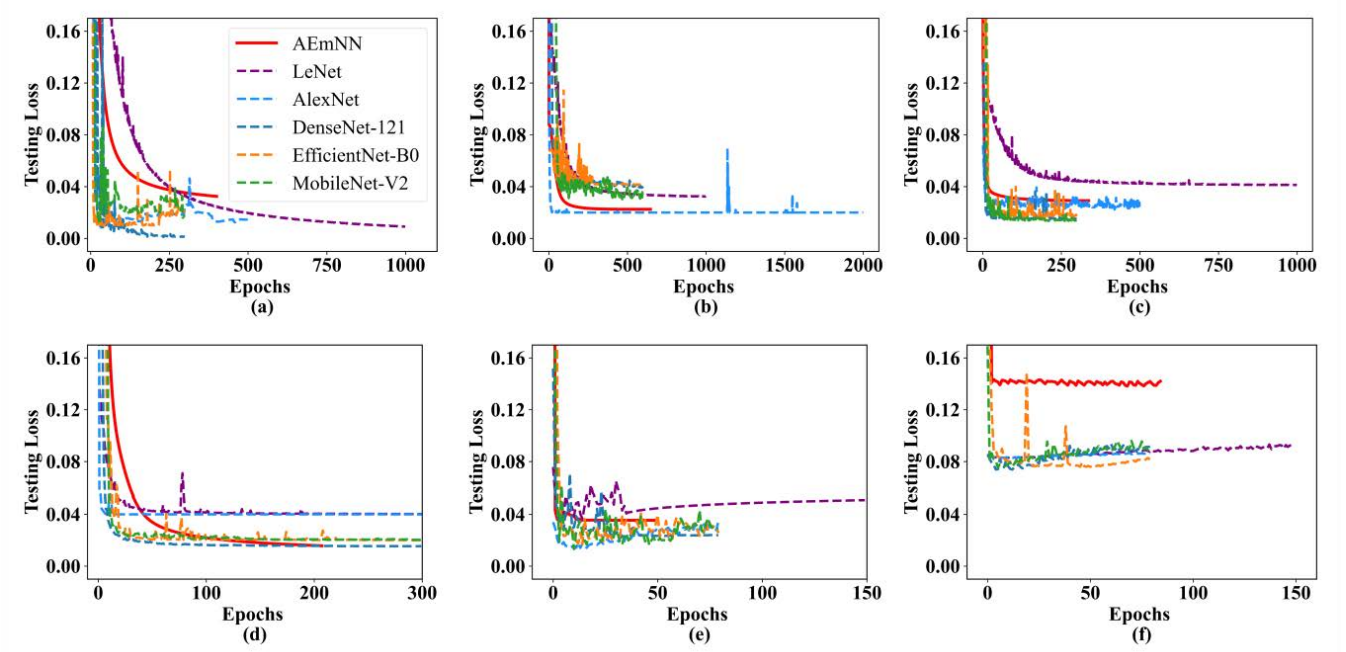


FIGURE 6. The testing loss curve of (a) ORL, (b)Yale, (c) Yale-B, (d) MIT, (e) MNIST, and (f) Fashion-MNIST fed to AEmNN, LeNet, AlexNet, DenseNet-121, EfficientNet-B0 and MobileNet-V2.

where y_m and t_m are the predicted and the target (label) of the m^{th} input pattern and p is the number of all patterns. AEmNN includes the emotional parameters of anxiety (μ_i) and confidence (k_i). The confidence parameter k is set to 0 at the start of training and gradually increases during training. The anxiety (μ_0) is obtained after the first epoch and then gradually decreases, and μ_i is the anxiety coefficient at the i^{th} epoch. These parameters are updated at each epoch according to the following rules:

$$\mu_i = y_{AvPAT} + rmse \tag{7}$$

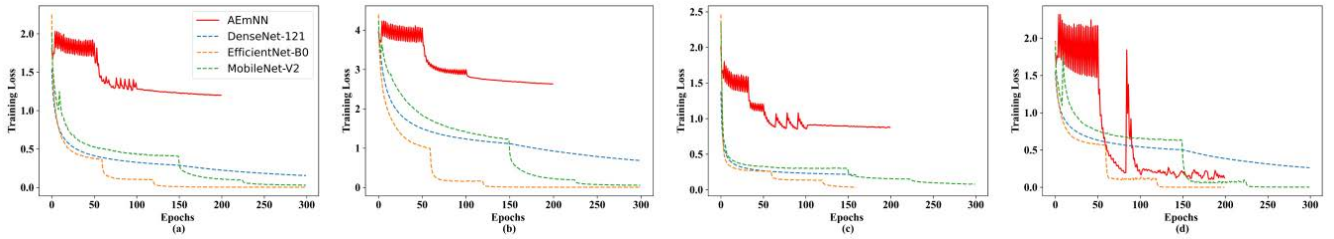


FIGURE 7. The learning curve of (a) CIFAR-10, (b) CIFAR-100, (c) SVHN, and (d) CINIC-10 fed to AEmNN, DenseNet-121, EfficientNet-B0 and MobileNet-V2.

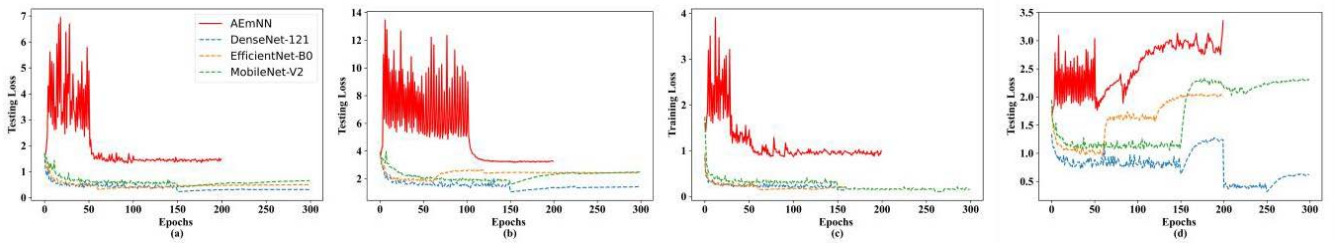


FIGURE 8. The testing loss curve of (a) CIFAR-10, (b) CIFAR-100, (c) SVHN, and (d) CINIC-10 fed to AEmNN, DenseNet-121, EfficientNet-B0 and MobileNet-V2.

$$k_i = |\mu_0 - \mu_i| \tag{8}$$

where y_{AvPAT} is determined by (8) and is the average value of p presented patterns to the network in each epoch, which is calculated as follows:

$$y_{AvPAT} = \frac{1}{p}\sum_{m=1}^{p} en_m \tag{9}$$

Here, we have proposed an adaptive emotional learning rate inspired from studies in [19], [24], [33], [40], [43], [44], [45], [46], and [48]. The value of the learning rate is adjusted by two controlling factors such as positive (im_{pos}) and negative (im_{neg}) impression. They are initialized adaptively at each epoch by decreasing and increasing value of anxiety and confidence parameters. Equations (10) and (11) illustrate their update method.

$$im_{pos} = exp(\mu_i) \tag{10}$$

$$im_{neg} = k_i \tag{11}$$

This procedure controls the adaptive value of learning rate based on the performance of the network, and positively affects the learning process and decision-making. Considering the widely accepted and proven evidence about the dopamine fluctuations behavior in humans/animals, we believe that the learning rate can mimic the dopamine fluctuations behavior according to emotional states (confidence and anxiety). The proposed adaptive emotional learning rate is determined as follows:

$$\eta_{i+1} = \begin{cases} \eta_i * im_{pos} & if\ e_i < e_{i-1} \\ \eta_i * im_{neg} & if\ e_i > e_{i-1} \\ \eta_i & if\ e_i = e_{i-1} \end{cases} \tag{12}$$

where η_i and η_{i+1} are the learning rate of i^{th} and $(i+1)^{th}$ epochs, e_{i-1} and e_i are the error of $(i\text{-}1)^{th}$ and i^{th} epochs. Equation (9), inspired form [51], which originally had two fixed decreasing and increasing factors, while in this study they changed into positive and negative impression factors. The value of im_{pos} is computed by the exponential value of anxiety (μ_i) in order to create a higher positive impression. This value updates the learning rate (η_{i+1}) when the i^{th} epoch is smaller than $(i\text{-}1)^{th}$ to help the network continue its path towards better learning and performance. Additionally, the value of im_{neg} is initialized by confidence (k_i) and update the learning rate (η_{i+1}) when the error in the epoch i is greater than (i-1). This process helps the network escape from probable local optima based on the network's confidence. Thus, this procedure determines the value of the learning rate based on the network's emotions and performance. In addition, it creates multiple fluctuations in learning rate during training, which leads to better and faster learning, similar to actual human dopamine fluctuations [20], [21], [22], [49], [52] or simulated dopamine fluctuation in Fig. 1 (b).

On the other hand, the value of the learning rate in fluctuations can reach infinity or zero. In order to solve this problem, the proposed method was inspired by studies in [20] and [53]. They did some experiments on humans/animals on different learning tasks and illustrated that dopamine fluctuates between two maximum and minimum values, which are different for each task.

Similar to these studies, we introduce a rule to control learning rate fluctuations. This rule allows the learning rate to fluctuate between the (η_{min}, η_{max}) range. Here, the lower (η_{min}) and upper (η_{max}) learning rate bound is initialized to 0 and 1 as it is often used in ANNs. The new value of the

TABLE 4. Training parameters and time convergence of each network for each dataset.

Dataset	Model	Training Time	Training Parameters	Train CE Loss
ORL	AEmNN	16 s	76 T	0.0017
	LeNet-5	20 s	64 T	0.0001
	AlexNet	142 s	3.9 M	0.0001
	EfficientNet-B0	160 s	4 M	0.0001
	MobileNet-V2	124 s	3.2 M	0.0001
	DenseNet-121	301 s	6.9 M	0.0001
Yale	AEmNN	4.9 s	32 T	0.003
	LeNet-5	7 s	64 T	0.001
	AlexNet	15 s	3.9 M	0.001
	EfficientNet-B0	17 s	4 M	0.001
	MobileNet-V2	12 s	3.2 M	0.001
	DenseNet-121	20 s	6.9 M	0.001
Yale-B	AEmNN	14 s	64 T	0.012
	LeNet-5	20 s	64 T	0.0009
	AlexNet	75 s	3.9 M	0.0002
	EfficientNet-B0	83 s	4 M	0.0003
	MobileNet-V2	61 s	3.2 M	0.0001
	DenseNet-121	135 s	6.9 M	0.0001
MIT	AEmNN	32 s	32 T	0.003
	LeNet-5	38 s	64 T	0.0001
	AlexNet	126 s	3.9 M	0.0001
	EfficientNet-B0	140 s	4 M	0.0001
	MobileNet-V2	94 s	3.2 M	0.0001
	DenseNet-121	421 s	6.9 M	0.0001
MNIST	AEmNN	343 s	156 T	0.1
	LeNet5	450 s	64 T	0.0001
	AlexNet	1090 s	3.9 M	0.0001
	EfficientNet-B0	1247 s	4 M	0.0001
	MobileNet-V2	392 s	3.2 M	0.0001
	DenseNet-121	1526 s	6.9 M	0.0001
Fashion-MNIST	AEmNN	502 s	156 T	0.15
	LeNet-5	498 s	64 T	0.0001
	AlexNet	1087 s	3.9 M	0.0001
	EfficientNet-B0	1383 s	4 M	0.0001
	MobileNet-V2	491 s	3.2 M	0.0001
	DenseNet-121	1713	6.9 M	0.0001
CIFAR-10	AEmNN	30 m	0.8 M	1.3
	EfficientNet-B0	1 h	4 M	0.007
	MobileNet-V2	30 m	3.2 M	0.01
	DenseNet-121	1.15 h	6.9 M	0.2
CIFAR-100	AEmNN	33 m	0.8 M	2.8
	EfficientNet-B0	1.30 h	4 M	0.001
	MobileNet-V2	1 h	3.2 M	0.003
	DenseNet-121	1.45 h	6.9 M	0.75
SVHN	AEmNN	20 m	0.8 M	0.89
	EfficientNet-B0	1.10 h	4 M	0.01
	MobileNet-V2	1 h	3.2 M	0.04
	DenseNet-121	1.30 h	6.9 M	0.3
CINIC-10	AEmNN	43 m	0.8 M	0.2
	EfficientNet-B0	2 h	4 M	0.001
	MobileNet-V2	1.45 h	3.2 M	0.001
	DenseNet-121	2.35 h	6.9 M	0.37

learning rate η_i is determined by (13) and is set to the initial learning rate η_{start} (initialized at the beginning of learning) if it is not in the assigned range.

$$\eta_i = \begin{cases} \eta_{start} & if\ \eta_i < \eta_{min} or \eta_i > \eta_{max} \\ \eta_i & if\ \eta_{min} \leq \eta_i \leq \eta_{max} \end{cases} \quad (13)$$

Moreover, the backpropagated error is denoted as δ_j for the output layer and δ_h for the hidden layer, which is computed

TABLE 5. Comparison of AEmNN and MLP, EmNN, PI-EmNN, LeNet5 and AlexNet in terms of train accuracy (%), test accuracy (%) and the number of epochs which was achieved in ten-time ten-fold cross-validation for each dataset. Train and test accuracy were reported as (mean ± std).

DataSet	Model	Train Acc %	Test Acc %	Epoch
ORL	**AEmNN**	**100 ± 0.0**	**98.5 ± 0.38**	**399**
	MLP	99.76 ± 0.08	97.07 ± 0.48	4850
	EmNN	99.89 ± 0.06	97.15 ± 0.49	6981
	PI-EmNN	99.91 ± 0.05	97.90 ± 0.43	2667
	LeNet-5	**100 ± 0.0**	98.45 ± 0.45	1000
	AlexNet	99.94 ± 0.04	98.21 ± 0.39	487
Yale	**AEmNN**	**100 ± 0.0**	**99.30 ± 0.21**	**646**
	MLP	98.88 ± 0.14	97.45 ± 0.78	3583
	EmNN	98.99 ± 0.13	97.53 ± 0.75	3068
	PI-EmNN	99.23 ± 0.10	98.80 ± 0.61	955
	LeNet-5	**100 ± 0.0**	98.64 ± 0.49	2000
	AlexNet	99.56 ± 0.12	99.03 ± 0.24	1000
Yale-B	**AEmNN**	99.54 ± 0.04	**99.20 ± 0.08**	**337**
	MLP	98.88 ± 0.03	97.45 ± 0.26	18777
	EmNN	98.99 ± 0.02	97.53 ± 0.26	16658
	PI-EmNN	99.23 ± 0.02	98.80 ± 0.17	2587
	LeNet-5	**100 ± 0.0**	98.64 ± 0.16	1000
	AlexNet	99.56 ± 0.19	99.03 ± 0.25	500
MIT	**AEmNN**	**99.99 ± 0.01**	**99.88 ± 0.02**	**207**
	MLP	98.48 ± 0.01	98.35 ± 0.11	15184
	EmNN	98.43 ± 0.03	98.20 ± 0.12	14980
	PI-EmNN	99.83 ± 0.01	99.70 ± 0.04	3019
	LeNet-5	100 ± 0.01	99.82 ± 0.04	300
	AlexNet	99.94 ± 0.04	99.79 ± 0.4	300
MNIST	**AEmNN**	99.72 ± 0.04	98.56 ± 0.15	**49**
	MLP	97.29 ± 0.14	96.25 ± 0.24	400
	EmNN	97.59 ± 0.14	96.56 ± 0.30	500
	PI-EmNN	98.49 ± 0.20	97.50 ± 0.25	200
	LeNet-5	99.03 ± 0.83	98.81 ± 0.82	80
	AlexNet	**99.90 ± 0.04**	**99.11 ± 0.13**	80
Fashion-MNIST	AEmNN	93.97 ± 0.53	90.50 ± 0.74	85
	MLP	90.01 ± 0.39	88.20 ± 0.48	100
	EmNN	90.31 ± 0.68	88.51 ± 0.52	500
	PI-EmNN	91.24 ± 0.33	89.90 ± 1.04	200
	LeNet-5	94.64 ± 0.92	90.23 ± 0.75	150
	AlexNet	**98.49 ± 0.39**	**90.76 ± 0.43**	**80**

by (14) and (15), respectively. Equations (14) and (15) are the final derivatives of the sigmoid activation function and cross-entropy loss function.

$$\delta_j = \left(t_j - y_j\right) \quad (14)$$

$$\delta_h = y_h \cdot (1 - y_h) \cdot \sum_{j=1}^{N_{classes}} w_{jh}.\delta_j \quad (15)$$

where y_h is the output value of neuron h in the hidden layer. In addition, the interconnection weights of the output layers and hidden layers are updated by (16-18) and (19-21), respectively:

$$w_{jh}\,(new) = w_{jh}\,(old) + \eta_i \cdot \delta_j \cdot y_h + \alpha \cdot [\delta w_{jh}(old)] \quad (16)$$

$$w_{jb}\,(new) = w_{jb}\,(old) + \eta_i \cdot \delta_j + \alpha \cdot [\delta w_{jb}(old)] \quad (17)$$

$$w_{je}\,(new) = w_{je}\,(old) + \mu_i \cdot \delta_j \cdot \mathrm{y}_{\mathrm{AvPAT}} + k_i \cdot [\delta w_{je}(old)] \quad (18)$$

TABLE 6. Comparison between AEmNN, EfficientNet-B0, MobileNet-V2 and DenseNet-121 in terms of train accuracy (%), test accuracy (%) and the number of epochs, which was achieved in ten independent runs for each dataset. Train and test accuracy were reported as (mean ± std).

DataSet	Method	Train Acc %	Test Acc %	Epoch
ORL	AEmNN	100 ± 0.0	98.5 ± 0.38	399
	EfficientNet-B0	100±0.0	97.31±0.029	300
	MobileNet-V2	100±0.0	97.25±0.024	300
	DenseNet-121	**100±0.0**	**100 ±0.0**	**300**
Yale	**AEmNN**	**100 ± 0.0**	**99.30 ± 0.21**	646
	EfficientNet-B0	99.27± 0.01	89.76 ± 0.11	600
	MobileNet-V2	100 ± 0.0	86.49 ± 0.09	600
	DenseNet-121	100 ± 0.0	84.40±0.08	600
Yale-B	**AEmNN**	**99.54 ± 0.04**	**99.20 ± 0.08**	337
	EfficientNet-B0	99.56±0.004	97.77± 0.02	300
	MobileNet-V2	99.99±0.01	98.56±0.01	300
	DenseNet-121	99.79±0.01	97.44± 0.02	300
MIT	AEmNN	99.99 ± 0.01	99.88 ± 0.02	207
	EfficientNet-B0	100 ± 0.0	99.89 ±0.01	300
	MobileNet-V2	100 ± 0.0	99.90 ±0.001	300
	DenseNet-121	**100 ± 0.0**	**99.97±0.01**	**300**
MNIST	AEmNN	99.72 ± 0.04	98.56 ± 0.15	**49**
	EfficientNet-B0	100 ± 0.0	99.01±0.01	80
	MobileNet-V2	100 ± 0.0	98.90±0.02	80
	DenseNet-121	**100 ± 0.0**	**99.42±0.01**	**80**
Fashion-MNIST	AEmNN	93.97 ± 0.53	90.50 ± 0.74	85
	EfficientNet-B0	96.17±0.02	88.18±0.01	80
	MobileNet-V2	99.60±0.01	90.01±0.02	80
	DenseNet-121	**99.71±0.02**	**91.74±0.02**	**80**
CIFAR-10	AEmNN	58.25±0.1	54.46±0.3	**200**
	EfficientNet-B0	99.97±0.02	91.01±0.2	300
	MobileNet-V2	99.52±0.04	87.67±0.11	300
	DenseNet-121	**99.93±0.05**	**93.02±0.24**	300
CIFAR-100	AEmNN	34.36±0.09	27.25±0.28	**200**
	EfficientNet-B0	99.96±0.02	64.23±0.7	300
	MobileNet-V2	83.48±0.05	59.26±0.94	300
	DenseNet-121	**85.63±0.61**	**71.04±0.47**	300
SVHN	AEmNN	75.54±0.19	75.10±0.30	200
	EfficientNet-B0	99.17±0.14	95.90±0.27	160
	MobileNet-V2	96.56±0.12	94.73±0.38	300
	DenseNet-121	**97.75±0.04**	**96.34±0.19**	**160**
CINIC-10	AEmNN	72.80±0.2	45.38±0.4	**200**
	EfficientNet-B0	94.98±0.24	72.41±0.59	200
	MobileNet-V2	92.06±0.30	67.64±0.51	300
	DenseNet-121	**96.74±0.21**	**79.64±0.38**	300

where i is the epoch index, $w_{jh}\,(new)$, $w_{jb}\,(new)$, $w_{je}\,(new)$, $\delta w_{jh}(old)$, $\delta w_{jb}(old)$, and $\delta w_{je}(old)$ are the updated weights and previous change in weights of the artificial neurons, bias and emotional neurons of interconnection weights of the output layers, respectively. Also, $w_{hi}\,(new)$, $w_{hb}\,(new)$, $w_{he}\,(new)$, $\delta w_{hi}(old)$, $\delta w_{hb}(old)$, and $\delta w_{he}(old)$ are the updated weights and previous change in weights of the artificial neurons, bias and emotional neurons of interconnection weights of the hidden layers, respectively. η_i is the adaptive emotional learning rate at epoch i and α is momentum.

$$w_{hi}\,(new) = w_{hi}\,(old) + \eta_i \cdot \delta_{\mathrm{h}} \cdot x_i + \alpha \cdot [\delta w_{hi}\,(old)] \quad (19)$$

$$w_{hb}\,(new) = w_{hb}\,(old) + \eta_i \cdot \delta_{\mathrm{h}} + \alpha \cdot \delta w_{hb}(old) \quad (20)$$

TABLE 7. The best available public results of modern deep networks on corresponding datasets. SVHN* with extra training data.

Dataset	Model	Best Test Accuracy %	Epoch
CIFAR-10	EfficientNet-B0[48]	98.1	90
	DenseNet-121 [46]	94.76	300
	MobileNet-V2 [47]	93.6	-
CIFAR-100	EfficientNet-B0[48]	88.1	90
	DenseNet-121 [46]	79.80	300
	MobileNet-V2 [47]	74.9	-
SVHN *	EfficientNet-B0	-	-
	DenseNet-121 [46]	98.21	40
	MobileNet-V2	-	-
CINIC-10	EfficientNet-B0	-	-
	DenseNet-121 [65]	91.26	-
	MobileNet-V2 [65]	82.0	-

$$w_{he}\,(new) = w_{he}\,(old) + \mu_i \delta_{\mathrm{h}} \cdot \mathrm{y}_{\mathrm{AvPAT}} + k_i \cdot [\delta w_{he}\,(\mathrm{old})] \quad (21)$$

III. DATASETS

This section introduces the employed datasets, which were used to evaluate the proposed method and its competitors in real-life tasks: ORL [54], Yale [55], [56], corrupt Yale-B [57], [58], MIT [59], [60], MNIST [61], Fashion-MNIST(F-MNIST) [62], CIFAR-10 [63], CIFAR-100 [63], SVHN [64] (without extra training data) and CINIC-10 [65]. Table 2 shows all the information concerned with the datasets. These datasets evaluate the proposed methods in different aspects, as described in Table 2. It is necessary to note that all input images are resized into 36*36 pixels for AEmNN and PI-EmNN networks to reduce computational requirements based on the work idea in [18]. Thus, each image is mapped into a 1024-dimensional vector which is denoted as input pattern $\mathbf{x}$.

IV. RESULTS AND DISCUSSION

This section presents the results of AEmNN and compares several techniques (MLP, EmNN, PI-EmNN, LeNet-5, AlexNet, DenseNet-121, MobileNet-V2 and EfficientNet-B0) through the ten-fold cross-validation for each dataset. The results are determined and demonstrated in terms of the convergence rate, accuracy, precision (22), recall (23), f1-score (24), r-squared (25), MAE (26), convergence rate, and two statistical tests. The convergence rate of the networks through the training phase was sketched in Fig. 3, Fig. 5 and Fig. 7. In addition, the testing loss of the networks were illustrated in Fig. 4, Fig. 6 and Fig. 8, demonstrating how an adaptive emotional learning rate (see Fig. 9 and Fig. 10), that mimics dopamine fluctuations, can speed up learning ability and help the proposed network make better decisions. To compare the network's accuracy, two statistical tests, Student's t-test [66] and Friedman's test [66] are used, which show that the proposed method has better or equal performance compared to its competitors.

$$precision = TP/(TP + FP) \quad (22)$$

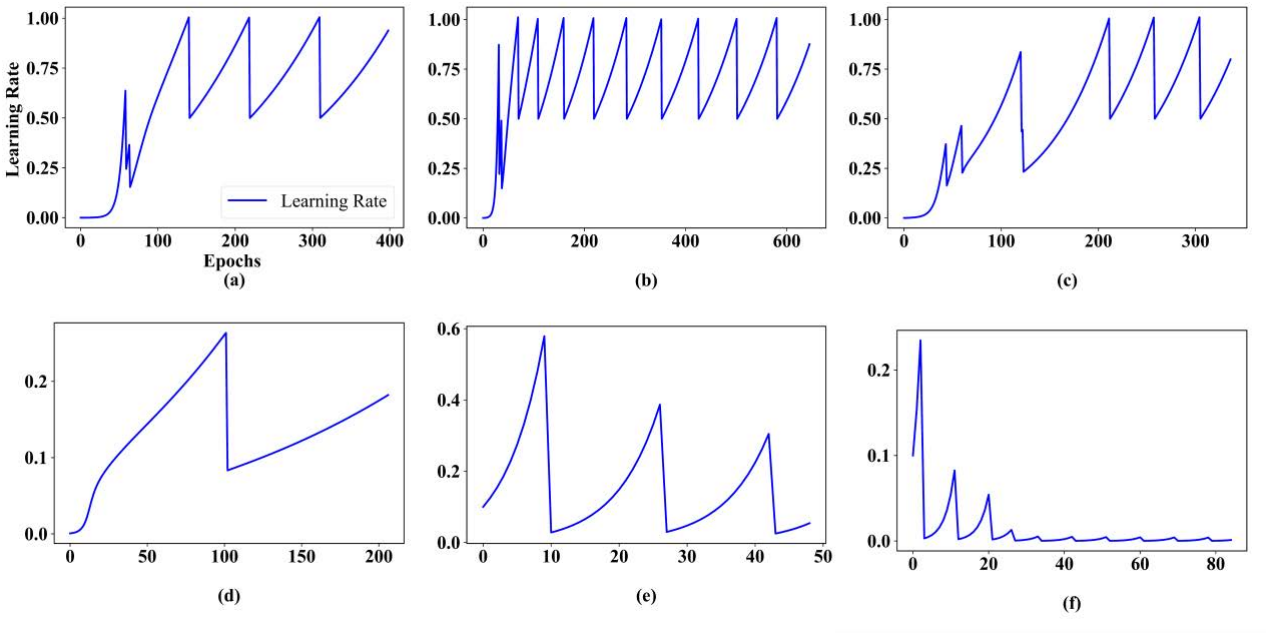


FIGURE 9. The curves of learning rate fluctuations (adaptive emotional learning rate) related to (a) ORL, (b) Yale, (c) Yale-B, (d) MIT, (e) MNIST, and (f) Fashion-MNIST in AEmNN.

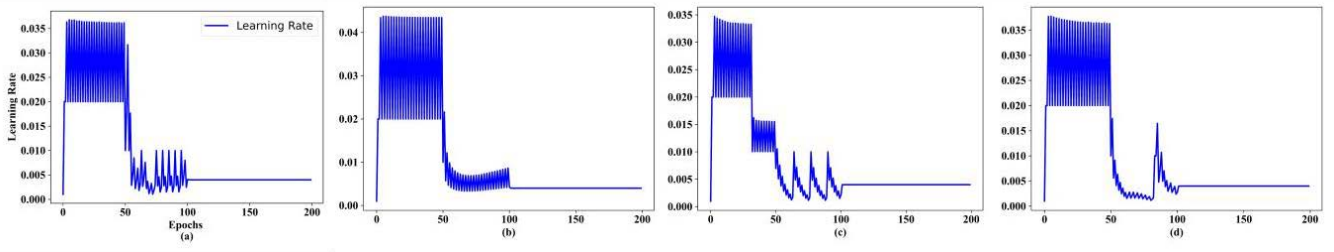


FIGURE 10. The curves of learning rate fluctuations (adaptive emotional learning rate) related to (a) CIFAR-10, (b) CIFAR-100, (c) SVHN, and (d) CINIC-10 in AEmNN.

$$recall = TP/(TP + FN) \tag{23}$$

$$f1 - score = (2 * precision * recall) /(precision + recall) \tag{24}$$

$$r - squared = 1 - \frac{\sum_{j=1}^{N_{classes}} (t_j - y_j)^2}{\sum_{j=1}^{N_{classes}} (t_j - \bar{t})^2} \tag{25}$$

$$mae = \frac{\sum_{j=1}^{N_{classes}} |t_j - y_j|}{N_{classes}} \tag{26}$$

where, TP, FN, and FP are true positives, false negatives, and false positives.

A. TRAINING SETTINGS

Training algorithms of each method were provided in Table 1, and parameter settings of all techniques, including the number of layers, initial learning rate, momentum, minimum error, training time, training parameters, mini-batch size and final anxiety/confidence values of emotional networks achieved during training are illustrated in Table 3. All parameters and weights for each model were fine-tuned through ten-fold internal cross-validation on a train set for datasets including ORL, Yale, corrupt Yale-B, MIT, MNIST and Fashion-MNIST. Besides, all parameters and settings of the deep networks such as LeNet-5, AlexNet, DenseNet-121, MobileNet-V2 and EfficientNet-B0 were set based on their own settings in their related article, and their learning rate and momentum were set to 0.01 and 0.9, respectively. In addition, Table 4 shows their number of training parameters and time convergence. It is important to note that all networks are implemented from scratch with PyTorch [67], Tensorflow [68] and Keras [69] libraries in python. Moreover, the proposed method, classical and emotional neural networks were executed by intel core i7 6th generation CPU, 16 GB RAM memory on windows 10 operating system. Also, deep networks were run with Kaggle website [70] providing 16 GB GPU memory (Nvidia P100) and 13 GB RAM memory.

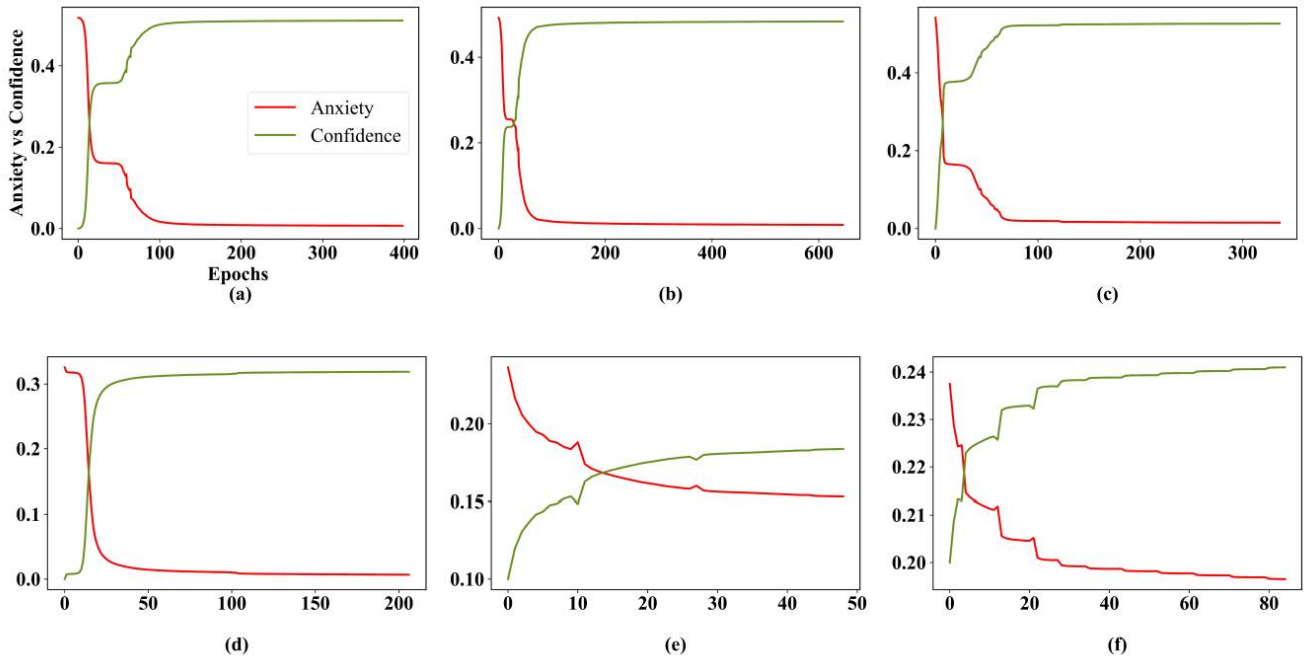


FIGURE 11. The learning curve of emotional parameters (anxiety coefficient (μ), and confidence coefficient (k)) of (a) ORL, (b) Yale, (c) Yale-B, (d) MIT, (e) MNIST, and (f) Fashion-MNIST in AEmNN.

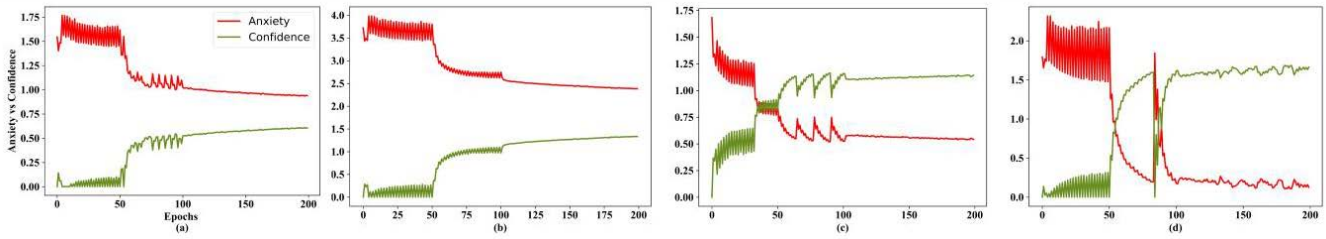


FIGURE 12. The learning curve of emotional parameters (anxiety coefficient (μ), and confidence coefficient (k)) of (a) CIFAR-10, (b) CIFAR-100, (c) SVHN, and (d) CINIC-10 in AEmNN.

B. PERFORMANCE AND CONVERGENCE RATE COMPARISON

The train accuracy, test accuracy, and the number of epochs in the training phase or the convergence rate of the compared methods applied to the described datasets are illustrated in Table 5 and Table 6. Besides, Table 7 illustrates the best available results of modern deep nets on CIFAR-10, CIFAR-100, SVHN, and CINIC-10. These results were obtained from their best run with data augmentation, which is not used in our study. In order to measure all models' performance in terms of precision, recall, f1-score, r-squared and MAE on all datasets, Table 8 and Table 9 were reported.

As can be seen, AEmNN has a better number of learning epochs, train and test accuracy and converges much faster than MLP, EmNN, PI-EmNN, LeNet-5 and AlexNet on ORL, Yale, Yale-B and MIT datasets. On the other hand, AEmNN is better than modern deep networks including EfficientNet-B0, DenseNet-121 and MobileNet-V2 in terms of accuracy and time converging time on Yale and Yale-B datasets. Based on mentioned terms, it is better than EfficientNet-B0, and MobileNet-V2 on ORL and Fashion-MNIST. On the contrary, it has lower performance on high-scale datasets such as CIFAR-10, CIFAR-100, SVHN and CINIC-10 compared to modern deep networks. In fact, AEmNN does not use any feature learning process or complicated structures like modern and classic deep networks. Besides, they have higher number of training parameters than AEmNN (please see Table 3 and Table 4). Fig. 3 and Fig. 4 show the proposed method's learning curve and testing loss curve compared to the MLP, EmNN, and PI-EmNN, and Fig. 5 and Fig. 6 show the proposed method's learning curve and testing loss curve compared to the LeNet-5, AlexNet, DenseNet-121, EfficientNet-B0 and MobileNet-V2 on ORL, Yale, Yale-B, MIT, MNIST and Fashion-MNIST. Besides, Fig. 7 and Fig. 8 show the proposed method's learning curve and testing loss curve compared to the DenseNet-121, EfficientNet-B0

TABLE 8. Compare AEmNN with MLP, EmNN, PI-EmNN, LeNet5 and AlexNet in terms of precision, recall, f1-score, r-squared and MAE, which was achieved from ten-time ten-fold cross-validation for train and test data of each dataset. Train and test results were reported as (mean ± std).

DataSet	Method	Train Precision %	Test Precision %	Train Recall %	Test Recall %	Train F1-Score %	Test F1-Score %	Train R-Squared	Test R-Squared	Train MAE	Test MAE
ORL	AEmNN	99.46±0.02	99.04±0.2	99.94±0.04	99.34±0.12	99.38±0.04	99.34±0.21	0.98±0.001	0.97±0.01	0.08±0.04	0.02±0.08
	MLP	100±0.0	97±0.3	100±0.0	98.0±0.5	100±0.0	97.33±0.9	1.0±0.0	0.98±0.02	0.08±0.02	0.3± 0.05
	EmNN	99.83±0.07	98.25±0.2	99.80±0.07	99.89±0.18	99.78±0.08	99.79±0.15	0.99±0.002	0.94±0.01	0.032±0.01	0.4± 0.06
	PI-EmNN	99.93±0.06	98.57±0.3	99.95±0.05	99.84±0.1	99.94±0.05	98.99±0.3	0.99±0.001	0.95±0.02	0.07±0.01	0.33±0.1
	LeNet-5	100 ± 0.0	98.51 ± 0.2	100±0.0	98 ± 0.51	100 ± 0.0	98.20 ±0.5	0.99±0.002	0.92 ± 0.05	0.0	0.01±0.002
	AlexNet	100 ± 0.0	98.15 ± 0.3	100±0.0	97.80 ± 0.23	100 ± 0.0	97.90 ±0.31	0.99±0.002	0.95 ± 0.05	0.0	0.03±0.004
Yale	AEmNN	99.64±0.01	99.64±0.1	100±0.0	99.60±0.2	99.76±0.02	99.60±0.1	0.98±0.02	0.94±0.06	0.03±0.02	0.08±0.02
	MLP	100±0.0	96±0.06	100±0.0	97.3±0.4	100±0.0	96.44±0.6	1±0.0	0.92±0.0	0.0	0.04±0.1
	EmNN	100±0.0	97.50±0.4	100±0.0	99.67±0.2	100±0.0	98.84±0.28	1.0±0.0	0.89±0.01	0.0	0.19±0.1
	PI-EmNN	100±0.0	99.23±0.2	100±0.0	100±0.0	100±0.0	99.49±0.2	1.0±0.0	0.97±0.01	0.0	0.08±0.2
	LeNet-5	100±0.0	100±0.0	100±0.0	93.75±0.02	100±0.0	96.79±0.15	1.0±0.0	0.98±0.02	0.0	0.001±0.01
	AlexNet	100±0.0	100±0.0	100±0.0	97.92±0.08	100±0.0	98.94±0.15	1.0±0.0	0.99±0.01	0.0	0.001±0.01
Yale-B	AEmNN	99.52±0.01	99.48±0.01	99.39±0.01	99.37±0.1	99.37±0.01	99.40±0.01	0.99±0.0	0.99±0.0	0.05±0.06	0.05±0.07
	MLP	99.71±0.01	98.64±0.2	99.69±0.02	97.80±0.3	99.69±0.01	97.81±0.4	0.99±0.0	0.98±0.04	0.03±0.02	0.3±0.1
	EmNN	99±0.01	98.12±0.01	98.88±0.01	97.52±0.01	98.90±0.01	97.42±0.02	0.97±0.01	0.94±0.04	0.18±0.05	0.34±0.2
	PI-EmNN	99.31±0.0	99.10±0.01	99.17±0.0	98.77±0.01	99.20±0.0	98.76±0.01	0.98±0.01	0.97±0.02	0.08±0.05	0.14±0.17
	LeNet-5	100±0.0	98.87 ± 0.01	100±0.0	98.15 ±0.01	100±0.0	98.31± 0.01	1.0±0.0	0.97±0.02	0.0	0.002 ± 0.0
	AlexNet	100±0.0	99.07 ± 0.07	100±0.0	98.95 ±0.04	100±0.0	99.08± 0.06	1.0±0.0	0.98±0.02	0.0	0.008 ± 0.0
MIT	AEmNN	99.92±0.001	99.90±0.001	99.92±0.001	99.89±0.001	99.92±0.001	99.89±0.001	0.99±0.001	0.99±0.001	0.001±0.01	0.002±0.01
	MLP	100±0.0	98.73±0.07	100±0.0	98.07±0.08	100±0.0	97.82±0.09	1.0±0.0	0.96±0.01	0.0	0.04±0.01
	EmNN	98.53±0.0	98.40±0.001	98.48±0.0	98.33±0.0.004	98.44±0.0	98.29±0.004	0.98±0.0	0.97±0.01	0.04±0.0	0.05±0.01
	PI-EmNN	99.86±0.0	99.72±0.002	99.86±0.0	99.72±0.002	99.86±0.0	99.72±0.002	0.99±0.001	0.99±0.005	0.004±0.001	0.007±0.007
	LeNet-5	100±0.0	99.85 ± 0.06	100±0.0	99.79 ±0.05	100±0.0	99.80± 0.04	1.0±0.0	0.96±0.03	0.0	0.001 ± 0.01
	AlexNet	100±0.0	99.78 ± 0.05	100±0.0	99.76 ±0.03	100±0.0	99.76± 0.02	1.0±0.0	0.96±0.03	0.0	0.001 ± 0.01
MNIST	AEmNN	99.63±0.08	98.45±0.24	99.64±0.05	98.52±0.31	99.63±0.01	98.50±0.21	0.99±0.01	0.96±0.02	0.02±0.002	0.06±0.02
	MLP	97.51 ±0.06	96.20±0.37	97.90±0.09	96.01±0.58	97.75±0.05	96.20±0.52	1.0±0.0	0.95±0.03	0.0	0.07±0.025
	EmNN	97.84 ±0.03	96.32±0.41	97.99±0.03	96.53±0.42	97.92±0.08	96.47±0.74	1.0±0.0	0.96±0.02	0.0	0.01±0.04
	PI-EmNN	98.32 ±0.07	97.60±0.84	98.90±0.09	97.40±0.58	98.60±0.02	97.49±0.62	1.0±0.0	0.98±0.01	0.0	0.01±0.05
	LeNet-5	99.05±0.05	98.89±0.2	98.99±0.03	98.57±0.21	99.01±0.015	98.72±0.24	1.0±0.0	0.99±0.001	0.0	0.002±0.001
	AlexNet	99.95±0.02	99.21±0.17	99.80±0.017	98.91±0.19	99.60.±0.013	99.05±0.24	1.0±0.0	0.99±0.001	0.0	0.001±0.001
Fashion-MNIST	AEmNN	93.68±0.02	89.44±0.3	93.69±0.04	89.64±0.25	93.55±0.01	89.32±0.32	0.89±0.01	0.83±0.27	0.2±0.05	0.3±0.17
	MLP	90.64 ± 0.07	89.01 ±0.45	90.50 ± 0.08	89.15 ±0.48	90.60 ± 0.04	89.01 ± 0.51	0.88 ± 0.06	0.82 ±0.48	0.3 ±0.08	0.35 ±0.28
	EmNN	90.41±0.04	88.30±0.2	90.86±0.05	88.42±0.6	90.60±0.01	88.33±0.39	0.85±0.05	0.81±0.31	0.24±0.02	0.41±0.23
	PI-EmNN	91.34±0.03	88.90±0.34	91.59±0.02	89.01±0.53	91.40±0.08	88.40±0.23	0.86±0.09	0.82±0.58	0.26±0.07	0.37±0.32
	LeNet-5	94.91± 0.02	89.56±0.32	94.20±0.004	88.90±0.24	97.51±0.004	89.24±0.41	0.96±0.01	0.8±0.1	0.01±0.002	0.02±0.12
	AlexNet	97.61± 0.07	90.20±0.41	97.41±0.004	90.41±0.45	97.51±0.004	90.30.±0.67	0.97±0.02	0.85±0.12	0.01±0.001	0.01±0.23

and MobileNet-V2 on CIFAR-10, CIFAR-100, SVHN, and CINIC-10. These figures depict that the learning curve of AEmNN is similar to the learning curve of classic and modern deep networks while the learning curves of AEmNN dropped neither fast nor slow because of using adaptive emotional learning rates, which guides learning and converges in a lower number of epochs in some datasets. In addition, the architectures of compared deep networks are complex, have more training parameters and need high computing resources to run, leading to a high convergence rate or the number of epochs.

Fig. 9 and Fig. 10 illustrate learning rate fluctuations of AEmNN on all datasets. These figures show that the learning rate in AEmNN can mimic human dopamine fluctuation behaviour based on studies in [19], [20], [22], [26], [36], and [37] and as shown in Fig. 1 (b). This process helps the network learn faster and better in some benchmark datasets.

The curves describing the learning of emotional parameters (anxiety (μ) and confidence coefficient (k)) of AEmNN for all datasets are illustrated in Fig. 11 and Fig. 12. It is observed that the anxiety coefficient is dropped as training progress in all figures, while the confidence coefficient is increased. Moreover, based on Table 3 AEmNN achieves the lowest anxiety coefficient (μ) and the highest confidence coefficient (k) compared to emotional networks (EmNN and PI-EmNN) in ORL, Yale, Yale-B, MIT, MNIST and Fashion-MNIST at the end of training.

C. STATISTICAL TESTS

To show that the proposed method is statistically better than other conventional and deep networks, the student t-test and Friedman test are applied to the network's test accuracy. These tests are used to show that the AEmNN accuracy is meaningfully better than the other compared networks or there is no meaningful difference between them.

TABLE 9. Compare AEmNN with EfficientNet-B0, MobileNet-V2 and DenseNet-121 in terms of precision, recall, f1-score, r-squared and MAE, which was achieved from ten independent runs for the train and test data of each dataset. Train and test results were reported as (mean ± std).

DataSet	Method	Train Precision %	Test Precision %	Train Recall %	Test Recall %	Train F1-Score %	Test F1-Score %	Train R-Squared	Test R-Squared	Train MAE	Test MAE
ORL	AEmNN	99.46±0.02	99.04±0.2	99.94±0.04	99.34±0.12	99.38±0.04	99.34±0.21	0.98±0.001	0.97±0.01	0.08±0.04	0.02±0.08
	EfficientNet-B0	100±0.0	97.73±0.03	100±0.0	97.25±0.03	100±0.0	96.41±0.01	1.0±0.0	0.95±0.05	0.0	0.002±0.001
	MobileNet-V2	100±0.0	98.00±0.03	100±0.0	97.00±0.02	100±0.0	96.33±0.03	1.0±0.0	0.96±0.03	0.0	0.002±0.001
	DenseNet-121	100±0.0	100±0.0	100±0.0	99.50±0.01	100±0.0	99.85±0.03	1.0±0.0	0.99±0.01	0.0	0.0
Yale	AEmNN	99.64±0.01	99.64±0.1	100±0.0	99.60±0.2	99.76±0.02	99.60±0.1	0.98±0.02	0.94±0.06	0.03±0.02	0.08±0.02
	EfficientNet-B0	99.26±0.01	96.16±0.08	99.26±0.01	86.84±0.1	99.25±0.01	86.86±0.2	0.98±0.02	0.85±0.13	0.001±0.0	0.02±0.01
	MobileNet-V2	100±0.0	92.60±0.08	100±0.0	82.91±0.09	100±0.0	83.15±0.25	1.0	0.79±0.1	0.0	0.02±0.01
	DenseNet-121	100±0.0	90.41± 0.1	100±0.0	82.55±0.08	100±0.0	80.01±0.07	1.0	0.77±0.12	0.0	0.02±0.0s1
Yale-B	AEmNN	99.52±0.01	99.48±0.01	99.39±0.01	99.37±0.1	99.37±0.01	99.40±0.01	0.99±0.0	0.99±0.0	0.05±0.06	0.05±0.07
	EfficientNet-B0	99.87± 0.01	98.33±0.01	99.47±0.01	97.60±0.02	99.57±0.02	97.76±0.02	0.99±0.01	0.97±0.03	0.0	0.002±0.001
	MobileNet-V2	100 ± 0.0	98.63±0.01	100 ± 0.0	98.51±0.01	100 ± 0.0	98.56±0.03	1.0	0.98 ± 0.02	0.0	0.0
	DenseNet-121	99.83±0.02	97.92±0.01	99.74±0.02	97.28±0.01	99.79±0.02	97.41±0.03	1.0	0.96 ± 0.03	0.0	0.02 ± 0.001
MIT	AEmNN	99.92±0.001	99.90±0.001	99.92±0.001	99.89±0.001	99.92±0.001	99.89±0.001	0.99±0.001	0.99±0.001	0.001±0.01	0.002±0.01
	EfficientNet-B0	100 ± 0.0	99.90±0.01	100 ± 0.0	99.89±0.01	100 ± 0.0	99.89±0.01	1.0	1.0	0.0	0.0
	MobileNet-V2	100 ± 0.0	99.91±0.01	100 ± 0.0	99.90±0.01	100 ± 0.0	99.90±0.01	1.0	1.0	0.0	0.0
	DenseNet-121	100 ± 0.0	99.97±0.02	100 ± 0.0	99.98±0.02	100 ± 0.0	99.97±0.01	1.0	1.0	0.0	0.0
MNIST	AEmNN	99.63±0.08	98.45±0.24	99.64±0.05	98.52±0.31	99.63±0.01	98.50±0.21	0.99±0.01	0.96±0.02	0.02±0.002	0.06±0.02
	EfficientNet-B0	100 ± 0.0	99.04±0.01	100 ± 0.0	99±0.01	100 ± 0.0	99.01±0.02	1.0	0.98±0.01	0.0	0.002±0.001
	MobileNet-V2	100 ± 0.0	98.93±0.02	100 ± 0.0	98.87±0.02	100 ± 0.0	98.89±0.01	1.0	0.97±0.02	0.0	0.002±0.001
	DenseNet-121	100 ± 0.0	99.43±0.01	100 ± 0.0	99.41±0.01	100 ± 0.0	99.42±0.01	1.0	0.98±0.01	0.0	0.001±0.01
Fashion-MNIST	AEmNN	93.68±0.02	89.44±0.3	93.69±0.04	89.64±0.25	93.55±0.01	89.32±0.32	0.89±0.01	0.83±0.27	0.2±0.05	0.3±0.17
	EfficientNet-B0	96.54±0.02	88.82±0.01	95.81±0.03	87.68±0.01	96.16±0.04	88.17±0.02	0.93±0.03	0.8±0.01	0.01±0.01	0.03±0.01
	MobileNet-V2	99.61±0.01	90.20±0.01	99.60±0.01	90.01±0.01	99.60±0.01	90.02±0.01	1.0	0.81±0.01	0.0	0.02±0.001
	DenseNet-121	99.70±0.01	91.81±0.02	99.71±0.01	91.68±0.02	99.70±0.01	91.70±0.02	1.0	0.84±0.01	0.0	0.02±0.001
CIFAR-10	AEmNN	58.69±0.2	54.19±0.37	58.26±0.24	54.42±0.35	58.44±0.22	54.29±0.26	0.06±0.02	0.0	1.49±0.34	1.68±0.48
	EfficientNet-B0	100 ± 0.0	90.98±0.12	100 ± 0.0	90.71±0.2	100 ± 0.0	90.86±0.1	1.0	0.82±0.01	0.0	0.3±0.14
	MobileNet-V2	99.49±0.03	87.64±0.41	99.50±0.02	87.01±0.3	99.49±0.04	87.30±0.26	1.0	0.74±0.05	0.01±0.02	0.4±0.11
	DenseNet-121	100 ± 0.0	93±0.35	100 ± 0.0	92.90±0.14	100 ± 0.0	92.93±0.33	1.0	0.85±0.03	0.0	0.24±0.24
CIFAR-100	AEmNN	27.49±0.08	20.13±0.51	28.41±0.11	21.59±0.49	27.96±0.14	20.87±0.42	0.0	0.0	22.35±2	25.15±3
	EfficientNet-B0	100 ± 0.0	55.04±0.7	100 ± 0.0	55.26±0.6	100 ± 0.0	53.12±0.5	1.0	0.23±0.2	0.0	12 ± 2
	MobileNet-V2	77.82±0.07	50.15±0.5	77.92±0.03	50.49±0.75	77.75±0.04	50.22±0.81	0.7±0.1	0.20±0.37	5.47±1.8	13.40±1.41
	DenseNet-121	85.63±0.06	62.45±0.49	80.59±0.09	62.69±0.72	79.04±0.08	62.47±0.75	0.72±0.24	0.43±0.24	4.78±1.33	9.49±1.94
SVHN	AEmNN	74.24±0.21	74.19±0.3	73.49±0.27	72.94±0.32	73.88±0.25	73.57±0.30	0.45±0.1	0.42±0.2	0.84±0.1	0.89±0.2
	EfficientNet-B0	99.24±0.22	95.72±0.39	99.09±0.31	95.54±0.47	99.15±0.25	95.60±0.35	0.98±0.01	0.89±0.02	0.03±0.01	0.15±0.05
	MobileNet-V2	96.48±0.17	94.43±0.58	96.35±0.26	94.41±0.42	96.39±0.21	94.42±0.29	0.92±0.01	0.87±0.03	0.11±0.02	0.2±0.041
	DenseNet-121	97.80±0.13	96.24±0.24	97.66±0.21	96.17±0.32	97.71±0.17	96.18±0.24	0.95±0.01	0.90±0.02	0.07±0.01	0.13±0.024
CINIC-10	AEmNN	72.76±0.19	45.02±0.41	72.77±0.24	45.30±0.44	72.76±0.21	45.17±0.40	0.4±0.09	0.0	0.94±0.1	1.98±0.2
	EfficientNet-B0	94.98±0.29	72.56±0.38	94.99±0.23	72.45±0.35	94.98±0.2	72.49±0.32	0.88±0.02	0.35±0.14	0.18±0.03	0.99±0.1
	MobileNet-V2	92.03±0.15	67.65±0.41	92.10±0.20	67.60±0.39	92.06±0.19	67.62±0.40	0.82±0.02	0.24±0.18	0.28±0.02	1.16±0.3
	DenseNet-121	96.75±0.21	79.69±0.35	96.66±0.19	79.64±0.31	96.73±0.20	79.65±0.33	0.93±0.01	0.50±0.12	0.11±0.01	0.74±0.15

TABLE 10. Training Student's *t*-test pairwise comparison in 95% confidence interval. *p-value* probability >*0.05* are indicated by bold-face.

Dataset	AEmNN vs							
	MLP	EmNN	PI-EmNN	LeNet-5	AlexNet	EfficientNet-B0	MobileNet-V2	DenseNet-121
ORL	0	0	0.01	0.1	0.04	0	0	0
Yale	0	0	0	0	0.01	0	0	0
Yale-B	0	0	0	0	**0.8**	0	0	0
MIT	0	0	0.009	**0.08**	0.04	0	0	0
MNIST	0	0	0	0.002	0	0	0	0
Fashion- MNIST	0	0	0	**0.07**	**0.5**	0	0	0
CIFAR-10	-	-	-	-	-	0	0	0
CIFAR-100	-	-	-	-	-	0	0	0
SVHN	-	-	-	-	-	0	0	0
CINIC-10	-	-	-	-	-	0	0	0

These tests compare the test accuracy of AEmNN with the other methods in pairs. They have null and alternative hypotheses. The null hypothesis is that the accuracy of AEmNN and the others is the same if the extracted p-value

TABLE 11. Training Friedman's test pairwise comparison in 95% confidence interval. *p-value* probability > *0.05* are indicated by bold-face.

Dataset	AEmNN vs							
	MLP	EmNN	PI-EmNN	LeNet-5	AlexNet	EfficientNet-B0	MobileNet-V2	DenseNet-121
ORL	0	0	0.002	0.6	0.03	0	0	0
Yale	0	0	0.003	0.01	0.04	0	0	0
Yale-B	0	0	0.001	0	**0.3**	0	0	0
MIT	0	0	0	**0.1**	0.02	0	0	0
MNIST	0	0	0	0.004	0	0	0	0
Fashion- MNIST	0	0	0	**0.06**	**0.09**	0	0	0
CIFAR-10	-	-	-	-	-	0	0	0
CIFAR-100	-	-	-	-	-	0	0	0
SVHN	-	-	-	-	-	0	0	0
CINIC-10	-	-	-	-	-	0	0	0

TABLE 12. Acronyms and abbreviations.

Acronym or Abbreviation	Full Name
ANN	Artificial Neural Network
EmNN	Emotional Neural Network
PI-EmNN	Prototype-Incorporated Emotional Neural Network
en	emotional neuron
CNN	Convolutional Neural Network
RPE	Reward Prediction Error
BP	Backpropagation
EmBP	Modified Emotional Backpropagation
SGD-adaptive	Stochastic Gradient Descent with adaptive emotional learning rate
BGD	Batch Gradient Descent
Mini-Batch SGD	Mini-Batch Stochastic Gradient Descent
CE Loss	Cross-Entropy Loss
MSE	Mean Squared Error
h, m, s	hour, minute, second
M, T	Million, Thousand
MLP	Multilayer Perceptron
Acc	Accuracy
MAE	Mean Absolute Error

probability is greater than the 0.05 significance level (95% confidence interval). In addition, the alternative hypothesis is that the accuracy of AEmNN is better than the others if the extracted *p-value* probability is lower than the 0.05 significance level.

T-test and Friedman test results were illustrated in Table 10 and Table 11 respectively. Based on the *t*-test and Friedman test results, AEmNN is not significantly better or worse than LeNet-5 in ORL, MIT, and Fashion-MNIST datasets. Still, it is significantly better than LeNet-5 in Yale, Yale-B, and MNIST datasets. It is noteworthy that AEmNN converges faster than LeNet-5 based on Table 3 and Table 5. In addition, AEmNN is not better than AlexNet in Yale-B, and Fashion-MNIST datasets, but AEmNN is significantly better than AlexNet in ORL, Yale, MIT, and MNIST.

Moreover, based on Table 10 and Table 11 AEmNN is significantly better than deep networks such as EfficientNet-B0 and MobileNet-V2 in ORL, Yale, Yale-B, and Fashion-MNIST datasets. It is significantly better than DenseNet-121 in Yale, and Yale-B datasets. On the other hand, DenseNet-121 is significantly better than AEmNN in datasets like CINIC-10, CIFAR-10, CIFAR-100, MIT, ORL and SVHN. It is noteworthy that AEmNN is significantly better than other techniques (MLP, EmNN, PI-EmNN) in all datasets.

V. CONCLUSION AND FUTURE WORK

In this research, we have proposed an efficient learning scheme for emotional neural networks and MLP, which directly and adaptively tunes the learning rate according to dopamine fluctuations in the human brain and emotions. The learning rate is updated adaptively at each epoch through stochastic learning, stabilizing the learning process. When the learning rate fluctuation is started, the role of dopamine fluctuation-inspired learning rate is highlighted, leading to the superiority of AEmNN to the compared techniques (which use a fixed learning coefficient or complex structures).

In most comparisons, the proposed method provides faster convergence, learning, better decision-making, and higher classification accuracy than compared techniques.

The proposed AEmNN outperforms state-of-the-art rivals like MLP, EmNN, PI-EmNN, LeNet-5, AlexNet, EfficientNet-B0, MobileNet-V2 and DenseNet-121 on different benchmarks in terms of the convergence rate and classification accuracy, but it has lower performance on high scale datasets compared to deep networks. It is noteworthy to restate that AEmNN has a simple (3 layers) architecture, and it does not use any feature learning process like deep networks. Simultaneously, its results are fairly comparable and, in some cases, better than LeNet-5, and AlexNet, DenseNet-121, EfficientNet-B0 and MobileNet-V2.

In addition to the dopamine hormone, other hormones like serotonin and adrenaline affect the decision-making and learning process of humans. Therefore, as a future work, investigating the behavior of other hormones and their learning role will be considered in emotional neural network's learning process or using AEmNN instead of the fully connected layer of deep networks or transfer learning networks in image classification tasks.

REFERENCES

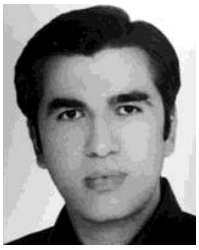
[1] C. M. Tyng, H. U. Amin, M. N. M. Saad, and A. S. Malik, "The influences of emotion on learning and memory," *Frontiers Psychol.*, vol. 8, p. 1454, Aug. 2017.

[2] A. Ben-Eliyahu, "Academic emotional learning: A critical component of self-regulated learning in the emotional learning cycle," *Educ. Psycholog.*, vol. 54, no. 2, pp. 84–105, Apr. 2019.
[3] D. Schuller and B. W. Schuller, "The age of artificial emotional intelligence," *Computer*, vol. 51, no. 9, pp. 38–46, Sep. 2018.
[4] S. Gu, M. Gao, Y. Yan, F. Wang, Y.-Y. Tang, and J. H. Huang, "The neural mechanism underlying cognitive and emotional processes in creativity," *Frontiers Psychol.*, vol. 9, p. 1924, Oct. 2018.
[5] A. Khashman, "Modeling cognitive and emotional processes: A novel neural network architecture," *Neural Netw.*, vol. 23, no. 10, pp. 1155–1163, 2010.
[6] S. E. Rivers, I. J. Handley-Miner, J. D. Mayer, and D. R. Caruso, "Emotional intelligence," in *The Cambridge Handbook of Intelligence*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2020, pp. 709–735.
[7] J. Chen, Y. Liu, and M. Zou, "User emotion for modeling retweeting behaviors," *Neural Netw.*, vol. 96, pp. 11–21, Dec. 2017.
[8] D. Levine, "Neural network modeling of emotion," *Phys. Life Rev.*, vol. 4, no. 1, pp. 37–63, Mar. 2007.
[9] L. I. Perlovsky, "Physics of the mind," *Frontiers Syst. Neurosci.*, vol. 10, p. 84, Nov. 2016.
[10] J. Yin, "Study on the progress of neural mechanism of positive emotions," *Transl. Neurosci.*, vol. 10, no. 1, pp. 93–98, Apr. 2019.
[11] R. Smith, W. D. S. Killgore, A. Alkozei, and R. D. Lane, "A neurocognitive process model of emotional intelligence," *Biol. Psychol.*, vol. 139, pp. 131–151, Nov. 2018.
[12] Y. Park, S. Baek, and S.-B. Paik, "A brain-inspired network architecture for cost-efficient object recognition in shallow hierarchical neural networks," *Neural Netw.*, vol. 134, pp. 76–85, Feb. 2021.
[13] R. Thenius, P. Zahadat, and T. Schmickl, "EMANN—A model of emotions in an artificial neural network," in *Proc. Adv. Artif. Life (ECAL)*, Sep. 2013, pp. 830–837.
[14] E. Lotfi and M.-R. Akbarzadeh-T, "Practical emotional neural networks," *Neural Netw.*, vol. 59, pp. 61–72, Nov. 2014.
[15] A. Khashman, "A modified backpropagation learning algorithm with added emotional coefficients," *IEEE Trans. Neural Netw.*, vol. 19, no. 11, pp. 1896–1909, Nov. 2008.
[16] A. G. Ranade, M. Patel, and A. Magare, "Emotion model for artificial intelligence and their applications," in *Proc. 5th Int. Conf. Parallel, Distrib. Grid Comput. (PDGC)*, Dec. 2018, pp. 335–339.
[17] E. Hudlicka, "To feel or not to feel: The role of affect in human–computer interaction," *Int. J. Hum.-Comput. Stud.*, vol. 59, nos. 1–2, pp. 1–32, Jul. 2003.
[18] O. K. Oyedotun and A. Khashman, "Prototype-incorporated emotional neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3560–3572, Aug. 2018.
[19] M. D. Humphries, J. A. Obeso, and J. K. Dreyer, "Insights into Parkinson's disease from computational models of the basal ganglia," *J. Neurol., Neurosurg. Psychiatry*, vol. 89, no. 11, pp. 1181–1188, Nov. 2018.
[20] K. T. Kishida, I. Saez, T. Lohrenz, M. R. Witcher, A. W. Laxton, S. B. Tatter, J. P. White, T. L. Ellis, P. E. Phillips, and P. R. Montague, "Subsecond dopamine fluctuations in human striatum encode superposed error signals about actual and counterfactual reward," *Proc. Natl. Acad. Sci. USA*, vol. 113, no. 1, pp. 200–205, Jan. 2016.
[21] D. Bang, K. T. Kishida, T. Lohrenz, J. P. White, A. W. Laxton, S. B. Tatter, S. M. Fleming, and P. R. Montague, "Sub-second dopamine and serotonin signaling in human striatum during perceptual decision-making," *Neuron*, vol. 108, no. 5, pp. 999–1010, 2020.
[22] R. J. Moran, K. T. Kishida, T. Lohrenz, I. Saez, A. W. Laxton, M. R. Witcher, S. B. Tatter, T. L. Ellis, P. E. Phillips, P. Dayan, and P. R. Montague, "The protective action encoding of serotonin transients in the human brain," *Neuropsychopharmacology*, vol. 43, no. 6, pp. 1425–1435, May 2018.
[23] A. A. Hamid, J. R. Pettibone, O. S. Mabrouk, V. L. Hetrick, R. Schmidt, C. M. V. Weele, R. T. Kennedy, B. J. Aragona, and J. D. Berke, "Mesolimbic dopamine signals the value of work," *Nature Neurosci.*, vol. 19, no. 1, pp. 117–126, Jan. 2016.
[24] A. Lak, M. Okun, M. M. Moss, H. Gurnani, K. Farrell, M. J. Wells, C. B. Reddy, A. Kepecs, K. D. Harris, and M. Carandini, "Dopaminergic and prefrontal basis of learning from sensory confidence and reward value," *Neuron*, vol. 105, no. 4, pp. 700–711, 2020.
[25] D. Wilson and T. R. Martinez, "The general inefficiency of batch training for gradient descent learning," *Neural Netw.*, vol. 16, no. 10, pp. 1429–1451, 2003.
[26] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, *Efficient BackProp BT–Neural networks: Tricks of the Trade*, 2nd ed., G. Montavon, G. B. Orr, K.-R. Müller, Eds. Berlin, Germany: Springer, 2012, pp. 9–48.
[27] S. J. Gershman and N. Uchida, "Believing in dopamine," *Nature Rev. Neurosci.*, vol. 20, no. 11, pp. 703–714, Nov. 2019.
[28] D. Mayer, E. Kahl, T. C. Uzuneser, and M. Fendt, "Role of the mesolimbic dopamine system in relief learning," *Neuropsychopharmacology*, vol. 43, no. 8, pp. 1651–1659, Jul. 2018.
[29] K. C. Berridge, "The debate over dopamine's role in reward: The case for incentive salience," *Psychopharmacology*, vol. 191, no. 3, pp. 391–431, Mar. 2007.
[30] M. Guitart-Masip, E. Duzel, R. Dolan, and P. Dayan, "Action versus valence in decision making," *Trends Cognit. Sci.*, vol. 18, no. 4, pp. 194–202, Apr. 2014.
[31] T. T.-J. Chong and M. Husain, "The role of dopamine in the pathophysiology and treatment of apathy," *Prog. Brain Res.*, vol. 229, pp. 389–426, Jan. 2016.
[32] J. D. Berke, "What does dopamine mean?" *Nature Neurosci.*, vol. 21, no. 6, pp. 787–793, Jun. 2018.
[33] M. Pessiglione, F. Vinckier, S. Bouret, J. Daunizeau, and R. Le Bouc, "Why not try harder? Computational approach to motivation deficits in neuro-psychiatric diseases," *Brain*, vol. 141, no. 3, pp. 629–650, Mar. 2018.
[34] W. Dabney, "A distributional code for value in dopamine-based reinforcement learning," *Nature*, vol. 577, pp. 671–675, Jan. 2020.
[35] H. Kasai, N. E. Ziv, H. Okazaki, S. Yagishita, and T. Toyoizumi, "Spine dynamics in the brain, mental disorders and artificial neural networks," *Nature Rev. Neurosci.*, vol. 22, no. 7, pp. 407–422, Jul. 2021.
[36] E. O. Neftci and B. B. Averbeck, "Reinforcement learning in artificial and biological systems," *Nature Mach. Intell.*, vol. 1, no. 3, pp. 133–143, Mar. 2019.
[37] A. Mohebi, J. R. Pettibone, A. A. Hamid, J.-M.-T. Wong, L. T. Vinson, T. Patriarchi, L. Tian, R. T. Kennedy, and J. D. Berke, "Dissociable dopamine dynamics for learning and motivation," *Nature*, vol. 570, no. 7759, pp. 65–70, Jun. 2019.
[38] C. Beste, N. Adelhöfer, K. Gohil, S. Passow, V. Roessner, and S.-C. Li, "Dopamine modulates the efficiency of sensory evidence accumulation during perceptual decision making," *Int. J. Neuropsychopharmacology*, vol. 21, no. 7, pp. 649–655, Jul. 2018.
[39] S. Cabib, C. Latagliata, and C. Orsini, "Role of stress-related dopamine transmission in building and maintaining a protective cognitive reserve," *Brain Sci.*, vol. 12, no. 2, p. 246, Feb. 2022.
[40] R. B. Rutledge, M. Moutoussis, P. Smittenaar, P. Zeidman, T. Taylor, L. Hrynkiewicz, J. Lam, N. Skandali, J. Z. Siegel, O. T. Ousdal, and G. Prabhu, "Association of neural and emotional impacts of reward prediction errors with major depression," *JAMA Psychiatry*, vol. 74, no. 8, pp. 790–797, Aug. 2017.
[41] D. Z. Lieberman and M. E. Long, *The Molecule of More: How a Single Chemical in Your Brain Drives Love, Sex, and Creativity-and Will Determine the Fate of the Human Race*. Dallas, TX, USA: BenBella Books, 2018.
[42] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, D. E. Rumelhart and J. L. McClelland, Eds. Cambridge, MA, USA: MIT Press, 1986, pp. 318–362.
[43] A. Khashman, "Application of an emotional neural network to facial recognition," *Neural Comput. Appl.*, vol. 18, no. 4, pp. 309–320, May 2009.
[44] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[45] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.
[46] G. Huang, Z. Liu, and K. Q. Weinberger, "Densely connected convolutional networks," 2016, *arXiv:1608.06993*.
[47] M. Sandler, A. G. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation," 2018, *arXiv:1801.04381*.
[48] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," 2019, *arXiv:1905.11946*.
[49] K. Chakroun, D. Mathar, A. Wiehler, F. Ganzer, and J. Peters, "Dopaminergic modulation of the exploration/exploitation trade-off in human decision-making," *eLife*, vol. 9, Jun. 2020, Art. no. e51260.

[50] P. A. Kragel, M. C. Reddan, K. S. LaBar, and T. D. Wager, "Emotion schemas are embedded in the human visual system," *Sci. Adv.*, vol. 5, no. 7, Jul. 2019, Art. no. eaaw4358.
[51] M. H. B. M. T. Hagan and H. B. Demuth, "Backpropagation," in *Neural Network Design*. Boston, MA, USA: PWS Publishing, 1996, pp. 131–206.
[52] B. Chew, T. U. Hauser, M. Papoutsi, J. Magerkurth, R. J. Dolan, and R. B. Rutledge, "Endogenous fluctuations in the dopaminergic midbrain drive behavioral choice variability," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 37, pp. 18732–18737, Sep. 2019.
[53] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with restarts," 2016, *arXiv:1608.03983*.
[54] A. L. Cambridge. (1994). *The ORL Database of Faces*. Cambridge, U.K. [Online]. Available: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[55] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
[56] (1997). *Yale Face Database*. [Online]. Available: http://vision.ucsd.edu/datasets/yale_face_dataset_original/yalefaces.zip
[57] K.-C. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, May 2005.
[58] (2005). *The Extended Yale Face Database B*. [Online]. Available: http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html
[59] MIT. (2004). *The MIT-CBCL Face Recognition Database*. Massachusetts Institute of Technology. [Online]. Available: http://cbcl.mit.edu/software-datasets/heisele/facerecognition-database.html
[60] B. Weyrauch, B. Heisele, J. Huang, and V. Blanz, "Component-based face recognition with 3D morphable models," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop*, Jun. 2004, p. 85.
[61] Y. LeCun. (1998). *The MNIST Database of Handwritten Digits*. [Online]. Available: http://yann.lecun.com/exdb/mnist
[62] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms," 2017, *arXiv:1708.07747*.
[63] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Univ. Toronto, Toronto, ON, Canada, Tech. Rep., 2009.
[64] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," in *Proc. NIPS Workshop*, 2011, pp. 1–9.
[65] L. N. Darlow, E. J. Crowley, A. Antoniou, and A. J. Storkey, "CINIC-10 is not ImageNet or CIFAR-10," 2018, *arXiv:1810.03505*.
[66] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Dec. 2006.
[67] A. Paszke, "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'single Alché-Buc, E. Fox, R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2019, pp. 8024–8035.
[68] M. Abadi, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*.
[69] F. Chollet, *Keras*. San Francisco, CA, USA: GitHub, 2015.
[70] Kaggle Team. (2021). *Kaggle*. [Online]. Available: https://www.kaggle.com/

MOHAMMAD AMIN ZARE was born in 1993. He received the B.Sc. degree in computer engineering from the Azad University of Marvdasht, Iran, in 2014, and the M.Sc. degree in computer engineering (artificial intelligence and robotics) from Shiraz University, Shiraz, Iran, in 2018. His current research interests include neural networks, deep learning, and statistical pattern recognition.

REZA BOOSTANI was born in 1973. He received the B.Sc. degree in electronics from Shiraz University, Shiraz, Iran, in 1996, and the M.Sc. and Ph.D. degrees in biomedical engineering from the Amirkabir University of Technology, Tehran, Iran, in 1999 and 2004, respectively. He has spent his research period with the Graz University of Technology in the BCI field, from 2002 to 2003. Since 2004, he has been a Faculty Member of the Computer Science and Engineering Department, Shiraz University. His current research interests include biomedical signal processing, statistical pattern recognition, and machine learning.



MOKHTAR MOHAMMADI received the B.S. degree in computer engineering from Shahed University, Tehran, Iran, in 2003, the M.S. degree in computer engineering from Shahid Beheshti University, Tehran, in 2012, and the Ph.D. degree in computer engineering from the Shahrood University of Technology, Shahrood, Iran, in 2018. He is currently with the Department of Information Technology, College of Engineering and Computer Science, Lebanese French University, Kurdistan Region, Iraq. His current research interests include signal processing, time-frequency analysis, and machine learning.

SAMANEH KOUCHAKI is currently a Lecturer in machine learning for healthcare at the Department of Electronic and Electrical Engineering and a member of the Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey. Her research interests include machine learning, health informatics, biomedical signal processing, and computational biology.