



Machine Learning Algorithms for High Performance Modelling in Health Monitoring System Based on 5G Networks

Juan Luis Meza Carhuancho

Department of Post Grade, Universidad Cèsar Vallejo, Perú, South America

Jacinto Joaquin Vertiz-Osores

Faculty of Engineering and Management, Universidad Nacional Tecnologica de Lima Sur, Perú

Maria Alina Cueva-Rios

Faculty of Engineering, Academic Department of Engineering of Food Industries, Universidad Nacional de Jaen, Perú

Doris Fuster-Guillen

Faculty of Education- National University of San Marcos, Perú

Yolvi Ocaña-Fernández

*Faculty of Education- National University of San Marcos, Perú
yocanaf@unmsm.edu.pe*

Zoila Mercedes Collantes Inga

Faculty of Systems Engineering and Informatics, Technological University of Perú, Perú

Article History	Abstract
<p>Received: 13 July 2022 Revised: 20 September 2022 Accepted: 26 October 2022</p>	<p>The development of Internet of Things (IoT) applications for creating behavioural and physiological monitoring methods, such as an IoT-based student healthcare monitoring system, has been accelerated by advances in sensor technology. Today, there are an increasing number of students living alone who are dispersed across large geographic areas, therefore it is important to monitor their health and function. This research propose novel technique in high performance modelling for health monitoring system by 5G network based machine learning analysis. Here the input is collected as EEG brain waves which are monitored and collected through 5G networks. This input EEG waves has been processed and obtained as fragments and noise removal is carried out. The processed EEG wave fragments has been extracted using K-adaptive reinforcement learning. this extracted features has been classified using naïve bayes gradient feed forward neural network. The performance analysis shows comparative analysis between proposed and existing technique in terms of accuracy, precision, recall, F-1 score, RMSE and MAP. Proposed technique attained accuracy of 95%, precision of 85%, recall of 79%, F-1 measure of 68%, RMSE of 52% and MAP of 66%.</p>
<p>CC License CC-BY-NC-SA 4.0</p>	<p>Keywords: <i>healthcare monitoring system, high performance modelling, 5G network, machine learning, EEG brain waves</i></p>

1. Introduction

A crucial role performed by nurses and doctors in ICU is patient monitoring. In truth, the intensive care unit (ICU) is a particular area in hospitals for patients with life-threatening illnesses and urgent conditions. These patients require the highest level of care from medical personnel, and they must be constantly monitored by monitoring systems [1]. The monitoring system is therefore regarded as a key

instrument. Due to digitisation, data have been rising dramatically in all fields in recent years. Massive data is referred to as "big data," and it cannot be processed by standard computers. Big data analytics is the process of analysing huge databases to find hidden patterns, new information, and value. Applications for big data analytics include healthcare, logistics delivery, fraud and risk detection, and weather forecasting [2]. We may investigate algorithms that employ enormous data sets to learn, generalise, and forecast with the use of ML algorithms. Making judgments and computational statistics are both directly related to ML. Applications for machine learning algorithms include forecasting product sales, determining the likelihood that rain will fall in a specific area, and many more. Medical professionals can identify illness patterns and determine the severity of the condition by conducting a systematic study of the available medical data. We can build prediction models for individualised treatment with aid of systematic analysis and ML methods. We can also keep track of patients' warning signs while they undergo clinical trials and help doctors choose the right medication for their patients [3]. IoT is a significant driver behind ICT technology development, guiding future industries toward automation and distributed intelligence. However, human body cannot simply be connected to Internet, unlike mechanical or digital technologies. A sensing and networking system can be used to connect a digital device to Internet [4]. We won't be able to connect human body to Internet even if it had a sensing system built into it. To conduct the measurement and determine the subject's health status, large sensing or measurement equipment is often used. Large measurement equipment, has a limitation in that it can only be utilised for a brief length of time under controlled settings. Therefore, with currently available enormous sensing and measuring technology, it is impossible to connect a person's body to Internet anywhere, at any time [5]. IoT applications in the areas of security and healthcare are severely constrained by inability to connect human body to Internet. Healthcare predictive analytics uses a variety of methodologies, including traditional linear models and cutting-edge AI and ML. DL, a branch of ML, is sufficiently trustworthy as well as resilient to automatically process and learn from a vast quantity of complex healthcare data and provides useful insights and answers to challenging issues. The outcomes of conventional models have been surpassed by its application to a wide range of medical applications. In particular, recurrent neural network (RNN) has gained prominence in study of temporal events with regard to time-sequential applications [6] and is proficient at managing long-term relationships of input data.

The contribution of this research is as follows:

1. To propose novel technique in high performance modelling for health monitoring system by 5G network based machine learning analysis
2. The processed EEG wave fragments has been extracted using K-adaptive reinforcement learning.
3. this extracted features has been classified using naïve bayes gradient feed forward neural network.

2. Related Works

Different techniques for predicting heart disease have been proposed recently. The use of several ensemble classifiers for improving heart disease risk prediction accuracy shows an accuracy of 85.4% [7]. An interval type-2 fuzzy logic method and rough sets-based attribute reduction using chaotic firefly method combine to produce an 86% accurate model for the diagnosis of heart disease [8]. A machine learning hybrid model that combines the linear method (LM) and random forest (RM) techniques to predict heart disease [9] has a performance accuracy of 88.7%. Using a fuzzy analytic hierarchy approach as well as an ANN for the tasks of feature weighting and classification, an integrated decision support system for forecasting risk of heart failure achieves 91.0% accuracy [10]. CDSS framework was proposed in study [11]. To forecast the presence and severity of chronic kidney disease (CKD), they used a deep neural network (DNN) classifier. A decision support system was presented by the author in [12] to examine the elderly's overall wellness at the Hong Kong community level. ML has received a lot of attention recently, particularly when it comes to identifying patterns in photos or unprocessed data. Work [13] discussed how advances in machine learning have made it possible for epidemiologists to sift through a large volume of digital data. The author [14] discussed the integration of social media platforms, ML and NLP to assist the study of large datasets for population-level mental health research. Some ML architecture stands out among various methodological variations. For instance, when classifying tweets into two classes—real instances of

allergy or awareness tweets—we found that the kNearest Neighbor (k-NN) classifier's precision was higher than that of various other ML classifiers, including NBM Modal, NB, and SVM [15]. In a similar vein, [16] shown that when compared to other classifiers like NB, RF and SVM, Multinomial NB Modal had the best text classification performance with an F-measure of 0. 811. To forecast the weekly status of the ILI-infected US population, the author [17] demonstrated a model using Twitter data and multilayer perceptron with backpropagation method. An integrated intelligent fuzzy expert system was built by the authors of [18] to forecast the course of renal failure. In this model, the GFR threshold value for renal failure prediction was set at 15 cc/kg/min/1.73m². SVM-based tools for forecasting the diabetic condition has been proposed in work [19]. They noted a 94% prediction accuracy in that trial [20].

3. System Model

This section discuss novel technique in high performance modelling for health monitoring system by 5G network based machine learning analysis. Here the input is collected as EEG brain waves which are monitored and collected through 5G networks. This input EEG waves has been processed and obtained as fragments and noise removal is carried out. The processed EEG wave fragments has been extracted using K-adaptive reinforcement learning. this extracted features has been classified using naïve bayes gradient feed forward neural network. Overall proposed architecture is given in figure-1.

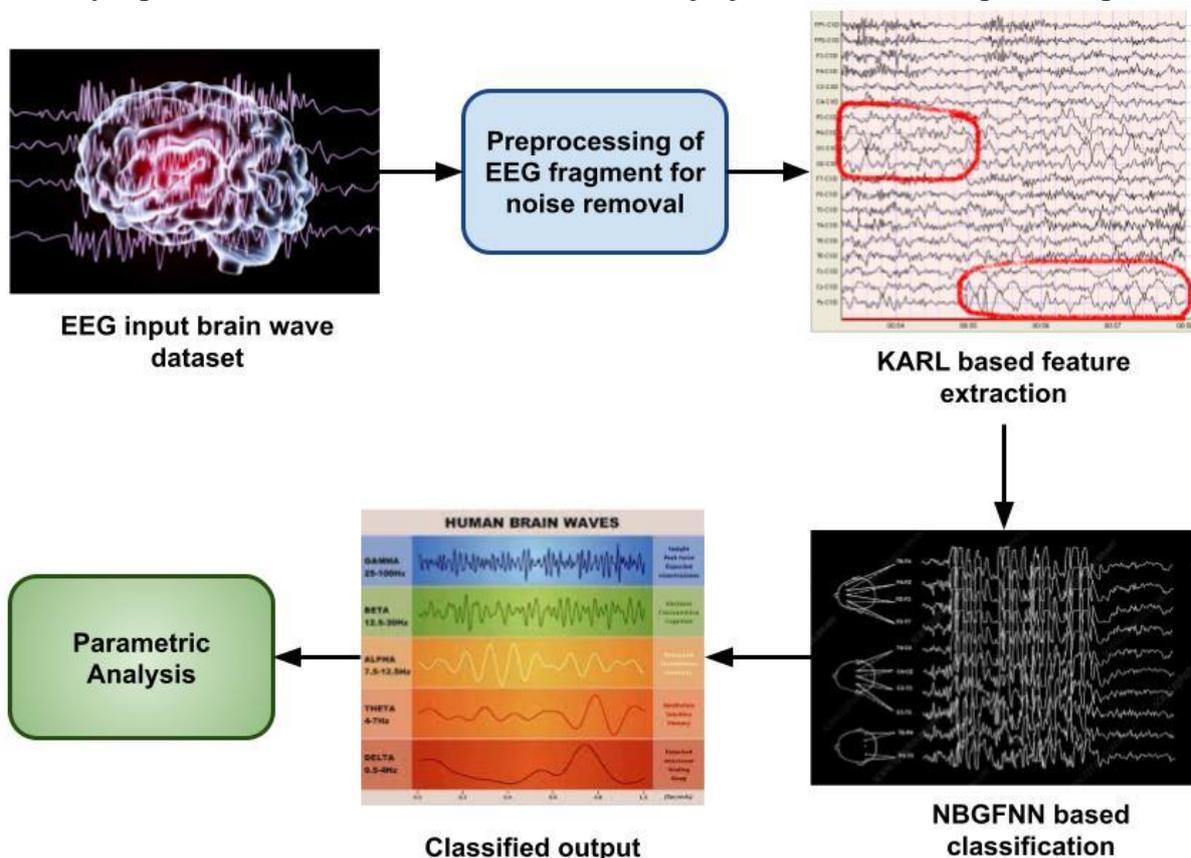


Figure 1. Overall proposed architecture

Pre-processing: To lessen the negative effects of signal abnormalities like crosstalk, noise, and power-line interference, data was pre-processed. The 10 centre sensors in this situation were eliminated from 64 EEG channels included in EEG test material due to their asymmetrical design. Then, two filters were applied to new 27 differential EEG channels: a notch filter eliminated interference from power lines that occurred at 50 Hz, and a band-pass filter was used to minimise artefacts like noise that are frequently present in this frequency range. The final step involved normalising EEG amplitude using min-max normalisation with various subjects to reduce the variation in EEG amplitudes.

3.1 K-adaptive reinforcement learning based feature extraction

Let's write $D = (x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)$ to represent training dataset for the desired task, where each tuple (x_i, y_i) represents each input image and its label in dataset. The d -dimensional parameter vector $\omega \in \mathbb{R}$ is then denoted as holding all d parameters of target method. Additionally, using transfer learning paradigms, one evaluate parameter of target network given a pretrained network with parameters based on a very large dataset as source. Goal of deep transfer learning-based optimization is to find equation(1) that minimises $L(\omega)$

$$\min_{\omega} \mathcal{L}(\omega) = \left\{ \frac{1}{n} \sum_{i=1}^n L(z(\mathbf{x}_i, \omega), y_h) + \lambda \cdot \Omega(\omega_1, \omega_n) \right\} \quad (1)$$

where (i) empirical loss of data fitting is described by first term, $\sum_{i=1}^n L(z(\mathbf{x}_i, \omega), y_i)$, and (ii) Second term $\Omega(\omega, \omega_s)$ describes variations in characteristics of the source and target networks. Trade-off between empirical loss as well as regularisation term is balanced by the tuning parameter $\lambda > 0$. $P(y_t|x_t)$, we used the maximum mean discrepancy (MMD), which is by equation (2):

$$D(D_s, D_t) \approx (1 - \mu) \left\| \frac{1}{n} \sum_{i=1}^n x_{si} - \frac{1}{m} \sum_{j=1}^m x_{tj} \right\|^2 + \mu \sum_{c=1}^C \left\| \frac{1}{n_c} \sum_{x_{sj} \in D_s} x_{si} - \frac{1}{m_c} \sum_{x_{tj} \in D_t} x_{tj} \right\|^2 \quad (2)$$

where H stands for RKHS, $c \in \{1, 2, \dots, C\}$ represents different class labels, n and m represent number of samples in source domain and target domain and D_s and D_t stand for samples that belong to class c in source domain and target domain. Number of samples in D_s and D_t is given by $n_c = |D_s|$ and $m_c = |D_t|$. Equation (3) is further applied with matrix techniques and regularisation to provide the following results:

$$\begin{aligned} \min_{\mathbf{A}} \text{tr} \left(\mathbf{A}^T \mathbf{X} \left((1 - \mu) \mathbf{M}_0 + \mu \sum_{c=1}^C \mathbf{M}_c \right) \mathbf{X}^T \mathbf{A} \right) + \lambda \|\mathbf{A}\|_F^2 \quad (3) \\ \text{s.t. } \mathbf{A}^T \mathbf{X} \mathbf{H} \mathbf{X}^T \mathbf{A} = \mathbf{I}, 0 \leq \mu \leq 1 \end{aligned}$$

Equation (4) has two terms: first term is regularisation term, and second term is the adaption of balance factor of conditional distribution and marginal distribution. In Equation, there are two restrictions (4). The inner attributes of the converted data (ATX) must be preserved while remaining compatible with original data. Second constraint places a range restriction on the balance factor where input data matrix X is made up of the elements x_s and x_t . A also stands for the transformation matrix by eq. (4):

$$\begin{aligned} (\mathbf{M}_0)_{ij} &= \begin{cases} \frac{1}{n^2}, & x_i, x_j \in D_s \\ \frac{1}{m^2}, & x_i, x_j \in D_t \\ -\frac{1}{mn}, & \text{otherwise} \end{cases} \\ (\mathbf{M}_c)_{ij} &= \begin{cases} \frac{1}{n_c^2}, & x_i, x_j \in D_s^{(c)} \\ \frac{1}{m_c^2}, & x_i, x_j \in D_t^{(c)} \\ \frac{-1}{m_c n_c}, & \begin{cases} x_i \in D_s^{(c)}, x_j \in D_t^{(c)} \\ x_i \in D_t^{(c)}, x_j \in D_s^{(c)} \end{cases} \\ 0, & \text{otherwise} \end{cases} \quad (4) \end{aligned}$$

After defining Lagrange multiplier as $\Phi = (\phi_1, \phi_2, \dots, \phi_d)$, Lagrange function for Equation (5) can be written as follows:

$$\begin{aligned} L = \text{tr} \left(\mathbf{A}^T \mathbf{X} \left((1 - \mu) \mathbf{M}_0 + \mu \sum_{c=1}^C \mathbf{M}_c \right) \mathbf{X}^T \mathbf{A} \right) \\ + \lambda \|\mathbf{A}\|_F^2 + \text{tr} \left((\mathbf{I} - \mathbf{A}^T \mathbf{X} \mathbf{H} \mathbf{X}^T \mathbf{A}) \Phi \right) \quad (5) \end{aligned}$$

Set derivative $\partial L / \partial \mathbf{A} = 0$ in this case. Then, one can transform the optimization of Equation (6) into a generalised eigen decomposition problem to arrive at:

$$\left(\mathbf{X} \left((1 - \mu) \mathbf{M}_0 + \mu \sum_{c=1}^C \mathbf{M}_c \right) \mathbf{X}^T \right) + \lambda \mathbf{I} = \mathbf{X} \mathbf{H} \mathbf{X}^T \mathbf{A} \Phi \quad (6)$$

The best transformation matrix A and its d lowest eigenvectors can be obtained simultaneously by solving Equation (6). The degree of similarity between tasks is a reliable sign of task consistency. The

gradients of similarity between an auxiliary task and the target goal as input by equation (9) are the weighting model's (sim ;) tasks.

$$\text{sim}_i = \cos(\nabla_{\theta} \mathcal{L}_{\text{sup}}(\theta), \nabla_{\theta} \mathcal{L}_{\text{aux},i}(\theta)) \quad (9)$$

where $\nabla_{\theta} \mathcal{L}_{\text{sup}}(\theta)$ is the gradient vector evaluated θ by the loss of auxiliary task ∇_{θ} and $\mathcal{L}_{\text{aux},i}(\theta)$ is the gradient vector evaluated at point by the loss of the target task. The joint loss can be expressed by the following equation, using (sim ;) to create weight for each task:

$$\underset{\theta}{\text{argmin}} g(\text{sim } m_0; w) \mathcal{L}_{\text{sup}}(\theta) + \sum_{i=1}^K g(\text{sim } m_i; w) \mathcal{L}_{\text{aux},i}(\theta) \quad (10)$$

We use the target task loss on the revised BP Adap TransL to optimise the weighting model after updating BP Adap TransL once by the aforesaid total loss. After one gradient update using the joint loss $L(\cdot)$, the revised parameters of the BP Adap TransL are indicated by the symbol θ_w .

We define the multi-task goal of BP Adap TransL as eq. (11): In the second stage, BP Adap TransL is trained by joint losses of both the target and auxiliary tasks. Given that the weights of various tasks are set by the weighting model.

$$\underset{\theta}{\text{argmin}} g(\text{sim } m_0; w) \mathcal{L}_{\text{sup}}(\theta) + \sum_{i=1}^K g(\text{sim } m_i; w) \mathcal{L}_{\text{aux},i}(\theta) \quad (11)$$

We repeatedly train both networks until they coincide. Agents that can perform a series of actions in an environment, represented by E , across a number of discrete time steps, denoted by $t \in \{1, \dots, T\}$, are taken into account by reinforcement learning. The agent creates a follow-up action at step S after receiving a state at step S at each time step t . After that, agent will see a new state, $S(t+1)$, and a scalar reward, r_t . The agents' objective is to discover a strategy $\pi(\mathbf{a}_t | \mathbf{s}_t)$ that maximises the objective function, which is the expected cumulative return with the form eq. (12)

$$J(\pi) := \sum_{t=0}^T \mathbb{E}_{(s_t, a_t)} [\gamma^t r(s_t, \mathbf{a}_t)] \quad (12)$$

where $0 \leq \gamma \leq 1$ is a discount rate, at is drawn from policy $\pi(\mathbf{a}_t | \mathbf{s}_t)$, and $\mathbf{s}_{t+1} = \mathcal{E}(\mathbf{s}_t, \mathbf{a}_t)$ is produced by running environment dynamics. The policy (a|s;) is parameterized by $\theta \in \mathbb{R}^d$ in policy-based reinforcement learning systems. where the parameters of the policy π , for example, the weights of a neural network, are represented by the vector $\theta := (\theta'_1, \dots, \theta'_d)^T$. The following optimization problem is then solved by eq(13) by iteratively changing the parameter θ after learning a good strategy,

$$\max_{\theta \in \mathbb{R}^d} J(\theta) \quad (13)$$

where we abuse notation to represent $J(\theta) := J(\pi(a|s; \theta))$. Automatic differentiation cannot be utilised to determine the gradient of $J(\theta)$ in reinforcement learning since it is typically true that the gradient of the environment S is inaccessible. Therefore, a large portion of advancement in reinforcement learning algorithms focuses on addressing the availability or absence of gradients in the environment or policy. A class of machine learning techniques known as reward shaping (RS) enables an agent to be trained on a synthetic reward signal as opposed to external feedback. However, there must be a potential role for this fake signal. If not, the new MDP will have a different optimal policy. If a real-valued function $\phi: S \rightarrow \mathbb{R}$ exists such that for all $s \in S$ by eq. (14), then F is a potential function.

$$\begin{aligned} F(s, s') &= \gamma \phi(s') - \phi(s) \\ F(s, a, s', a') &= \gamma \phi(s', a') - \phi(s, a) \\ F(s, t, s', t') &= \gamma \phi(s', t') - \phi(s, t) \\ F(s, a, t, s', a', t') &= \gamma \phi(s', a', t') - \phi(s, a, t) \end{aligned} \quad (14)$$

Equation 14 will be produced by utilising the Q-learning update rule with reward shaping. Any of the equations shown in equation 14 could be F . Equation 14 will give us the chance to improve learning by giving the Q-learning agent some additional knowledge about the issue. This additional information may originate from any source, such as human problem-solving expertise, a heuristic algorithm, or information that the agent has learned through artificial intelligence eq.(15).

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [R_{t+1} + F + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)] \quad (15)$$

γ is unchanged and must have the same value as it had in the previous MDP in equation 3 and all of its extended equations. Additionally, numerous literary works exist that expand the function's potential.

$$\eta(\pi) = \mathbb{E}_{s_0, a, \dots} [\sum_{t=0}^{\infty} \gamma^t r(s_t)], \text{ where } s_0 \sim \rho_0(s_0), a_t \sim \pi(a_t | s_t), s_{t+1} \sim P(s_{t+1} | s_t, a_t). \quad (16)$$

We shall employ the conventional definitions of the advantage function A by equation (17), value function V_{π} and state action value function Q_{π} as follows:

$$\begin{aligned} Q_{\pi}(s_t, a_t) &= \mathbb{E}_{s_{t+1}, a_{t+1}, \dots} [\sum_{l=0}^{\infty} \gamma^l r(s_{t+l})] \\ V_{\pi}(s_t) &= \mathbb{E}_{a_t, s_{t+1}, \dots} [\sum_{l=0}^{\infty} \gamma^l r(s_{t+l})] \\ A_{\pi}(s, a) &= Q_{\pi}(s, a) - V_{\pi}(s), \text{ where} \end{aligned} \quad (17)$$

$a_t \sim \pi(a_t | s_t), s_{t+1} \sim P(s_{t+1} | s_t, a_t)$ for $t \geq 0$

The following important identity describes how the advantage π over another strategy, expressed in terms of expected return, is built over timesteps by equation (18):

$$\eta(\tilde{\pi}) = \eta(\pi) + \mathbb{E}_{s_0, a_0, \dots \sim \tilde{\pi}} [\sum_{t=0}^{\infty} \gamma^t A_{\pi}(s_t, a_t)] \quad (18)$$

where notation $\mathbb{E}_{s_0, a_0, \dots \sim \pi} [\dots]$ is actions are sampled at $a_t \sim \tilde{\pi}(\cdot | s_t)$. Let ρ_{π} be discounted visitation frequencies by eq. (19)

$$\rho_{\pi}(s) = P(s_0 = s) + \gamma P(s_1 = s) + \gamma^2 P(s_2 = s) + \dots, \quad (19)$$

where $s_0 \sim \rho_0$ and the actions are selected using π . Equation (1) can be rewritten as follows using a sum over states rather than timesteps:

$$\begin{aligned} \eta(\tilde{\pi}) &= \eta(\pi) + \sum_{t=0}^{\infty} \sum_s P(s_t = s | \tilde{\pi}) \sum_a \tilde{\pi}(a | s) \gamma^t A_{\pi}(s, a) \\ &= \eta(\pi) + \sum_s \sum_{t=0}^{\infty} \gamma^t P(s_t = s | \tilde{\pi}) \sum_a \tilde{\pi}(a | s) A_{\pi}(s, a) \\ &= \eta(\pi) + \sum_s \rho_{\tilde{\pi}}(s) \sum_a \tilde{\pi}(a | s) A_{\pi}(s, a) \end{aligned} \quad (20)$$

This equation implies that every policy update $\pi \rightarrow \tilde{\pi}$ that has an expected advantage that is nonnegative at every state s , i.e., $\sum_a \tilde{\pi}(a | s) A_{\pi}(s, a) > 0$, is guaranteed to raise the policy performance η . The expected advantage will normally be negative for some states s in the approximate situation, i.e. $\sum_a \tilde{\pi}(a | s) A_{\pi}(s, a) < 0$, however this is usually unavoidable due to estimating and approximation error. Instead, we present the local approximation to as follows:

$$L_{\pi}(\tilde{\pi}) = \eta(\pi) + \sum_s \rho_{\pi}(s) \sum_a \tilde{\pi}(a | s) A_{\pi}(s, a) \quad (21)$$

L_{π} ignores variations in state visiting density brought on by changes in policy and instead uses visitation frequency ρ_{π} instead of $\rho_{\tilde{\pi}}$ visitation density. But if we have a parameterized policy π_{θ} , L_{π} matches to η first order if $\pi_{\theta}(a|s)$ is a differentiable function of the parameter vector. For every parameter value that is θ_0 by eq. (22),

$$\begin{aligned} L_{\pi_{\theta_0}}(\pi_{\theta_0}) &= \eta(\pi_{\theta_0}) \\ \nabla_{\theta} L_{\pi_{\theta_0}}(\pi_{\theta}) \Big|_{\theta=\theta_0} &= \nabla_{\theta} \eta(\pi_{\theta}) \Big|_{\theta=\theta_0} \end{aligned} \quad (22)$$

3.2 Naive bayes gradient feed forward neural network based classification

Let A and B represent two instances from sample space Ω , which can either be countably infinite or finite with N elements. Event where both A and B happen is indicated by $A \cap B$, or the intersection of events A and B. Finally, two events, A and B, are said to be mutually exclusive if their occurrence eliminates the likelihood that the other will also occur. This indicates that A and B are disjoint in set theory notation, i.e., $A \cap B = \emptyset$.

$$\begin{aligned} (A | B) &= \frac{|A \cap B|}{|B|} = \frac{\frac{|A \cap B|}{|M|}}{\frac{|B|}{|M|}} = \frac{P(A \cap B)}{P(B)} \\ P(B | A) &= \frac{|B \cap A|}{|A|} = \frac{\frac{|B \cap A|}{|M|}}{\frac{|A|}{|M|}} = \frac{P(A \cap B)}{P(A)} \end{aligned} \quad (23)$$

From Eq. 23, it is immediately obvious that eq. (24)

$$P(A | B) = \frac{P(B|A)P(A)}{P(B)} \quad (24)$$

which is the Bayes theorem's most straightforward (and possibly most famous) statement. The generalised Bayes' formula is eq. (25) and if B is an event with $P(B) > 0$, which is a subset of the union of all A_i

$$P(A_i | B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)} \quad (25)$$

which can be rewritten as eq. (26)

$$P(A | B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} \quad (26)$$

It is frequently believed that the data might emerge from two opposing hypotheses, H1 and H2, with $P(H1) = 1 - P(H2)$. It is also common to use the word "model" for the word "hypothesis." Let D stand for the measured data. Then, eq(27) provides posterior probability of hypothesis H1

$$P(H_1 | D) = \frac{P(D|H_1)P(H_1)}{P(D|H_1)P(H_1) + P(D|H_2)P(H_2)} \quad (27)$$

and posterior probability of H2 is by eq. (28), (29)

$$P(H_2 | D) = \frac{P(D|H_2)P(H_2)}{P(D|H_1)P(H_1)+P(D|H_2)P(H_2)} \quad (28)$$

$$\frac{P(H_1|D)}{P(H_2|D)} = \frac{P(D|H_1)}{P(D|H_2)} \cdot \frac{P(H_1)}{P(H_2)} \quad (29)$$

posterior odds Baye factor B Bi2 prior odds

Ratio of H1's posterior odds to its prior odds is known as Bayes factor. When compared to the competing hypothesis H2, Bayes factor is seen as a summary indicator of evidence data give us supporting hypothesis H1. Bayes factor and posterior odds are the same if the prior probability for H1 and H2 are the same. Let's define their conditional probability density functions as $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$. Next, by eq (30)

$$f_{X|Y}(x | y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

$$f_{Y|X}(y | x) = \frac{f_{XY}(x,y)}{f_X(x)} \quad (30)$$

so that eq. (31) can be used to express Bayes' theorem for continuous variables

$$f_{X|Y}(x | y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)} \quad (31)$$

where $f_Y(y) = \int_x f_{Y|X}(y | x)f_X(x)dx = \int_{-\infty}^{+\infty} f_{XY}(x, y)dx$ because of total probability theorem. These probabilities are used by the straightforward naive Bayes classifier to categorise an instance. By using Bayes' theorem and slightly condensing notation, we arrive at by eq (32)

$$P(y_j | \mathbf{x}_i) = \frac{P(\mathbf{x}_i|y_j)P(y_j)}{P(\mathbf{x}_i)} \quad (32)$$

Therefore, numerator is recast as follows; here, we'll just use x and leave off index I for convenience by eq. (33):

$$\begin{aligned} P(\mathbf{x} | y_j)P(y_j) &= P(\mathbf{x}, y_j) \\ &= P(x_1, x_2, \dots, x_p, y_j) \\ &= P(x_1 | x_2, x_3, \dots, x_p, y_j)P(x_2, x_3, \dots, x_p, y_j) \text{ because } P(a, b) = P(a | b)P(b) \\ &= P(x_1 | x_2, x_3, \dots, x_p, y_j)P(x_2 | x_3, x_4, \dots, x_p, y_j)P(x_3, x_4, \dots, x_p, y_j) \\ &= P(x_1 | x_2, x_3, \dots, x_p, y_j)P(x_2 | x_3, x_4, \dots, x_p, y_j) \cdots P(x_p | y_j)P(y_j) \end{aligned} \quad (33)$$

Assume for the moment that each xi is distinct from the others. Strong and obviously broken in the majority of practical applications, this assumption is naïve, hence the term. According to this presumption, by eq (34)

$$P(x_1 | y_j) \cdot P(x_2 | y_j) \cdots P(x_p | y_j)P(y_j) = \prod_{k=1}^p P(x_k | y_j)P(y_j) \quad (34)$$

which we can plug into Eq. 34 and we obtain by eq. (35)

$$P(y_j | \mathbf{x}) = \frac{\prod_{k=1}^p P(x_k|y_j)P(y_j)}{P(\mathbf{x})} \quad (35)$$

The denominator, P(x), is independent of the class; it is the same for classes yj and yl, for instance. The scaling factor P(x) makes sure that posterior probability P(yj | x) is appropriately scaled. The maximal posterior rule is the name given to this rule. The maximal a posteriori (MAP) class, which is the resulting "winning" class, is determined as y for instance x as eq. (36):

$$\hat{y} = \operatorname{argmax}_{y_j} \prod_{k=1}^p P(x_k | y_j)P(y_j) \quad (36)$$

A basic naive Bayes classifier is a model that applies Eq. 21. However, a clear categorisation is not always preferred. These natural ranking scores represent the estimated class posterior probabilities. Eq. 37 can be rewritten as, using total probability theorem (Eq. 3), where

$$P(y_j | \mathbf{x}) = \frac{\prod_{k=1}^p P(x_k|y_j)P(y_j)}{\prod_{k=1}^p P(x_k|y_j)P(y_j)+\prod_{k=1}^p P(x_k|y_j^c)P(y_j^c)} \quad (37)$$

We will give an example of how to determine the general form using eq(38) approximate answer using our method

$$x^m y''(x) = F(x, y(x), y'(x)) \quad (38)$$

where the domain is denoted by m, $m \in \mathbb{Z}$, $x \in \mathbb{R}$, $D \subset \mathbb{R}$ represents the answer that needs to be computed. Eq. (39) converts the issue to a discretize form if (x, p) signifies a trial solution with movable parameters p:

$$\text{Min}_p \sum_{x_i \in \bar{D}} F(x_i, y_t(x_i, p), y'_t(x_i, p)) \quad (39)$$

confining by the BCs' restrictions, of course. For the trial function (x), we select a form that fulfils the BCs. This is done by expressing it as the product of two terms in equation (40):

$$y_t(x_i, p) = A(x) + G(x, N(x, p)) \quad (40)$$

When the single-output FFNN $N(x, p)$ is given the input vector x along with the parameters p and n . The term (x) fulfils the BCs and has no changeable parameters. Since (x) satisfies BCs, second term G is designed to not add to them. This term can be created by employing an FFNN whose biases and weights have been modified to address the minimization issue. The FFNN can be trained using an effective minimization of (3), where error associated with each input x_i is value that must be forced close to zero. This error value is calculated using both the FFNN output and its derivatives with regard to any of its inputs.

For a given input x , output of FFNN is given by eq. (41)

$$N = \sum_{i=1}^l y_i \sigma(z_i), \text{ where } z_i = \sum_{j=1}^n w_{ij} x_j + b_i \quad (41)$$

It is simple to obtain the gradient of the FFNN with regard to its parameters as an equation (42)

$$\begin{aligned} \frac{\partial N}{\partial v_i} &= \sigma(z_i) \\ \frac{\partial N}{\partial b_i} &= v_i \sigma'(z_i) \\ \frac{\partial N}{\partial w_{ij}} &= v_i \sigma'(z_i) \end{aligned} \quad (42)$$

Additionally, it should be noted that weight adjustments may be performed in batch mode. We will use by eq(43) as an example to show how the method works

$$\frac{x^{m'} d^2 y(x)}{dx^2} = f(x, y, y') \quad (43)$$

where $x \in [a, b]$ and the BC: $y(a) = A, y(b) = B$; a trial solution is written as eq. (44)

$$y_t(x, p) = \frac{(bA-aB)}{(b-a)} + \frac{(B-A)x}{(b-a)} + (x-a)(x-b)N(x, p) \quad (44)$$

where $N(x, p)$ represents result of an FFNN with P weights and one unit of input for x . You should be aware that $y(x)$ meets BC by construction. eq(45) provides error quantity that needs to be minimised

$$E[p] = \sum_{i=1}^n \left\{ \frac{d^2 y_t(x_i, p)}{dx^2} - f\left(x_i, y_t(x_i, p), \frac{dy_t(x_i, p)}{dx}\right) \right\}^2 \quad (45)$$

where the $x_i \in [a, b]$. Since by eq. (46)

$$\begin{aligned} \frac{dy_t(x, p)}{dx} &= \frac{(B-A)}{(b-a)} + \{(x-a) + (x-b)\}N(x, p) \\ &\quad + (x-a)(x-b) \frac{dN(x, \vec{p})}{dx} \\ \frac{d^2 y_t(x, p)}{dx^2} &= 2N(x, p) + 2\{(x-a) + (x-b)\} \frac{dN(x, \vec{p})}{dx} \\ &\quad + (x-a)(x-b) \frac{d^2 N(x, p)}{dx^2} \end{aligned} \quad (46)$$

Gradient of error with regard to specification p is easily calculated using (46).

4. Performance Analysis

4.1 Dataset Description

This study made use of the GAMEEMO dataset, which is openly accessible. Signals gathered from participants playing emotional video games are included in the EEG dataset. To compare neurological signals and meaningful responses across different people and EEG sessions, participants were asked to complete trials for 4 sessions apiece. These EEG signals were obtained while playing video games. They were collected utilising a 14-channel wearable and portable electroencephalography (EEG) instrument on an Emotive EPOC+ from 28 different individuals. There was a total of 20 minutes of EEG data available for each subject. Subjects engaged in four different emotional computer games for five minutes each. This dataset aims to evaluate the efficacy of wearable EEG devices to traditional EEG devices and to provide an alternate source of data for the emotion identification process.

DEAP is a multimodal dataset for the study of emotional states in people. 32 volunteers viewed 40 one-minute highlight music videos while their EEG and peripheral physiological data were monitored. Suggested ensemble model's findings were assessed in our experiments using eight-fold cross-

validation, and outcomes are shown as an average standard deviation. All of the studies had a total of 30 iterations. Table 1 displays the many steps in the EEG signal processing.

Table 1. Proposed technique-based EEG signal processing for various dataset

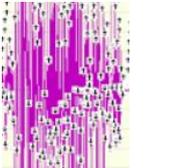
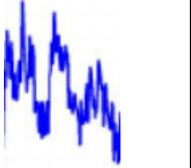
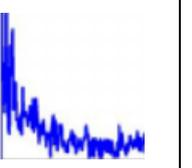
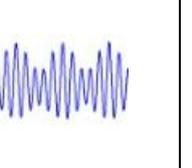
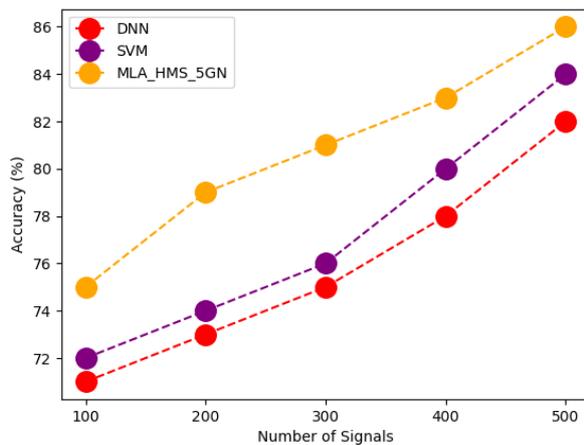
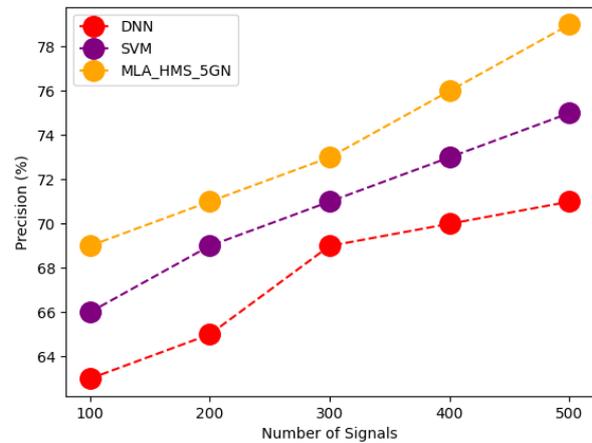
Dataset	Input signal	EEG	Processed input signal	Extracted features of EEG signal	Classified EEG signal	Detected emotion from classified signal
EEG Database dataset						
SSVEP						

Table 2. Comparative analysis between proposed and existing technique for various dataset

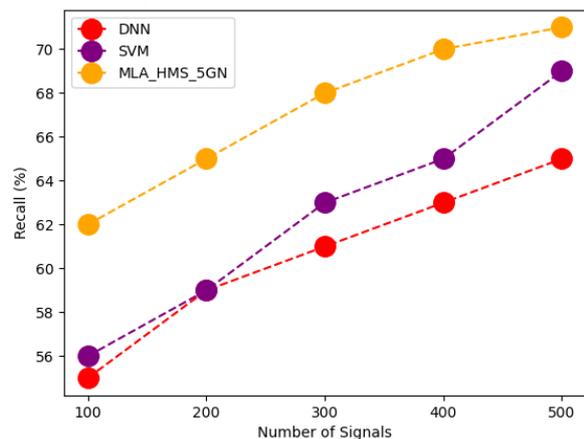
Datasets	Techniques	Accuracy	Precision	Recall	F1_Score	RMSE	MAP
GAMEEMO	DNN	82	71	65	55	41	52
	SVM	84	75	69	59	43	55
	MLA_HMS_5GN	86	79	71	62	45	59
DEAP	DNN	85	81	68	59	45	61
	SVM	90	83	75	65	49	63
	MLA_HMS_5GN	95	85	79	68	52	66



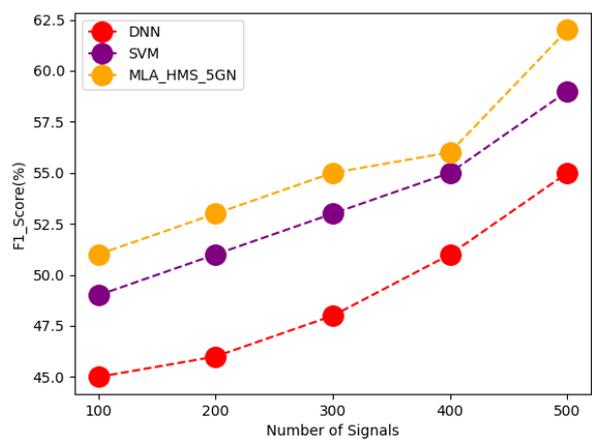
(a) Accuracy



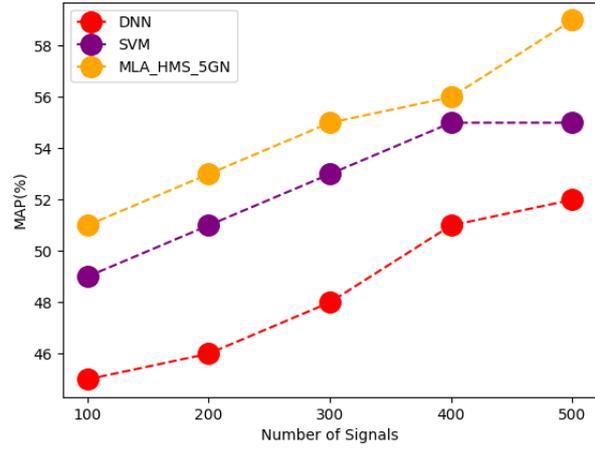
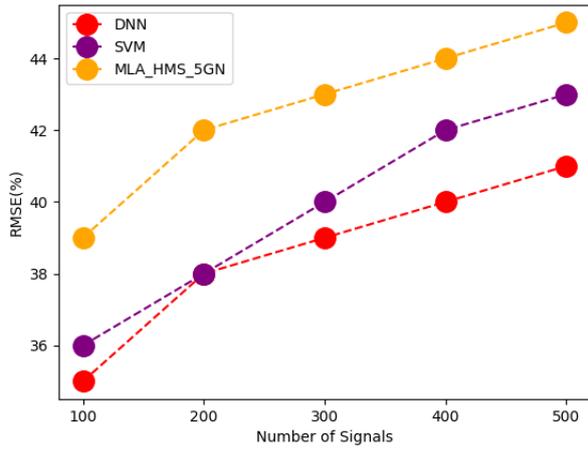
(b) Precision



(c) Recall



(d) F-1 score

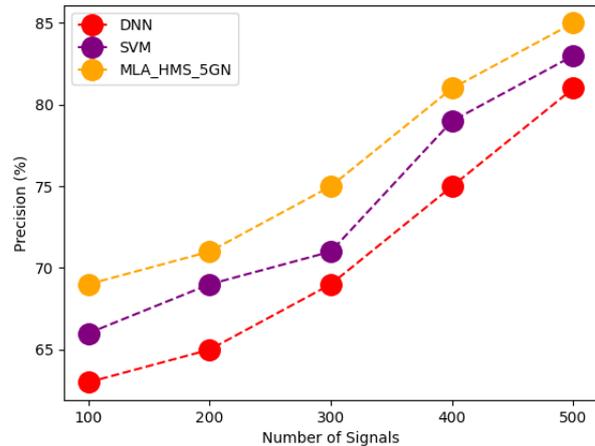
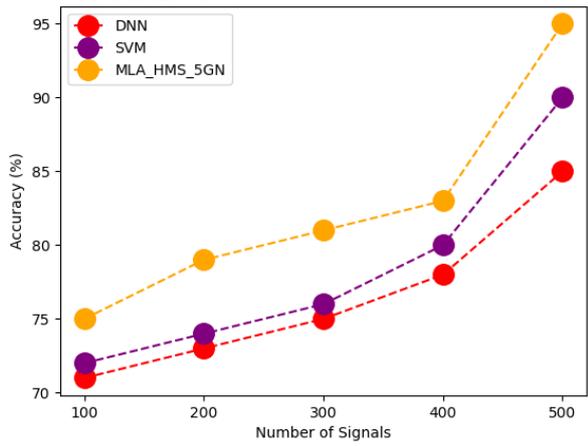


(e) RMSE

(f) MAP

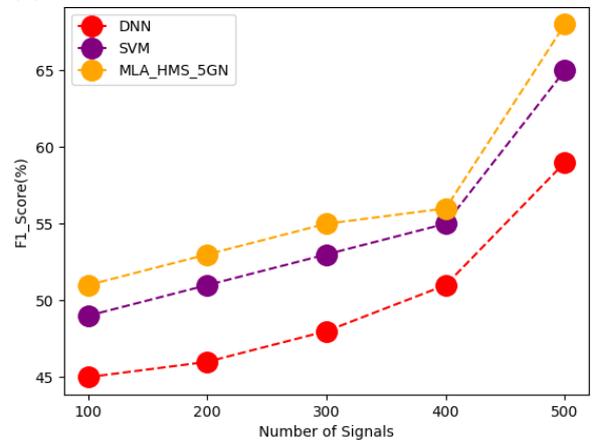
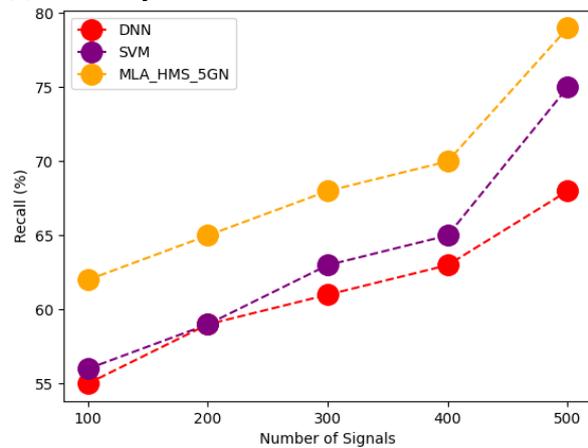
Figure 2. Comparative analysis between proposed and existing technique for GAMEEMO dataset in terms of (a) Accuracy, (b) precision, (c) recall, (d) F-1 score, (e) RMSE, (f) MAP

The above figure-2 shows comparative analysis between proposed and existing approach for GAMEEMO dataset. Here proposed technique attained accuracy of 86%, precision of 79%, recall of 71%, F-1 measure of 62%, RMSE of 45% and MAP of 59%, existing technique DNN attained accuracy of 82%, precision of 71%, recall of 65%, F-1 measure of 55%, RMSE of 41% and MAP of 52%, SVM attained accuracy of 84%, precision of 75%, recall of 69%, F-1 measure of 59%, RMSE of 43% and MAP of 55%.



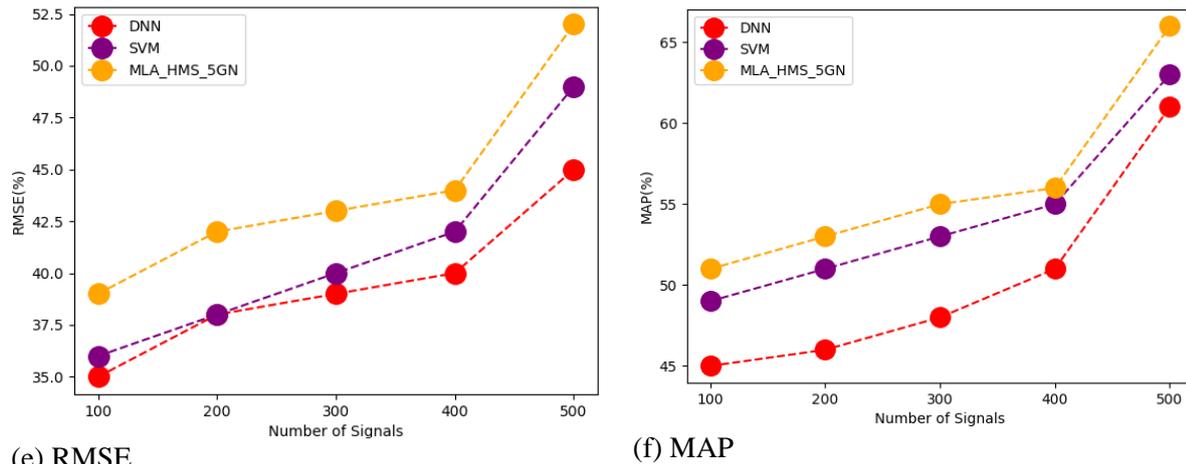
(a) Accuracy

(b) Precision



(c) Recall

(d) F-1 score



(e) RMSE (f) MAP
 Figure 3. Comparative analysis between proposed and existing technique for DEAP dataset in terms of (a) Accuracy, (b) precision, (c) recall, (d) F-1 score, (e) RMSE, (f) MAP

From above figure-3 the comparative analysis between proposed and existing technique for DEAP dataset. Proposed approach attained accuracy of 95%, precision of 85%, recall of 79%, F-1 measure of 68%, RMSE of 52% and MAP of 66%, existing technique DNN attained accuracy of 85%, precision of 81%, recall of 68%, F-1 measure of 59%, RMSE of 45% and MAP of 61%, SVM attained accuracy of 90%, precision of 83%, recall of 75%, F-1 measure of 65%, RMSE of 49% and MAP of 63%.

5. Conclusion

The proposed framework has been designed novel technique in health monitoring system by 5G network using machine learning analysis. The processed EEG wave fragments has been extracted using K-adaptive reinforcement learning. this extracted features has been classified using naïve bayes gradient feed forward neural network. Performance of propose network was calculated by evaluating classification accuracy, varying different parameters such as number of strides, learning rate parameter, number of epochs, and sample size. An extensive data learning approach is proposed to classify depression EEG signals from that of healthy controls. The key advantage of using deep learning is that they return state-of-the-art accuracy and do not require manual pre-processing or feature extraction from the signal. Proposed approach attained accuracy of 95%, precision of 85%, recall of 79%, F-1 measure of 68%, RMSE of 52% and MAP of 66%. Instead of categorization using deep neural networks, future research will concentrate on methods capable of early detection or prediction of hand movements. Moreover, the utilisation of hand-crafted features could be a drawback of our research.

References

- [1] Berrar, D. (2018). Bayes' theorem and naive Bayes classifier. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 403.
- [2] Tawfiq, L. N., & Hussein, A. A. (2013). Design feed forward neural network to solve singular boundary value problems. *International Scholarly Research Notices*, 2013.
- [3] AlShorman, O., Masadeh, M., Heyat, M. B. B., Akhtar, F., Almahasneh, H., Ashraf, G. M., & Alexiou, A. (2022). Frontal lobe real-time EEG analysis using machine learning techniques for mental stress detection. *Journal of Integrative Neuroscience*, 21(1), 20.
- [4] Bressan, G., Cisotto, G., Müller-Putz, G. R., & Wriessnegger, S. C. (2021). Deep learning-based classification of fine hand movements from low frequency EEG. *Future Internet*, 13(5), 103.
- [5] Guerrero, M. C., Parada, J. S., & Espitia, H. E. (2021). EEG signal analysis using classification techniques: Logistic regression, artificial neural networks, support vector machines, and convolutional neural networks. *Heliyon*, 7(6), e07258.

- [6] Topic, A., & Russo, M. (2021). Emotion recognition based on EEG feature maps through deep learning network. *Engineering Science and Technology, an International Journal*, 24(6), 1442-1454.
- [7] Sun, J., Cao, R., Zhou, M., Hussain, W., Wang, B., Xue, J., & Xiang, J. (2021). A hybrid deep neural network for classification of schizophrenia using EEG Data. *Scientific Reports*, 11(1), 1-16.
- [8] Zhang, R., Jia, J., & Zhang, R. (2022). EEG analysis of Parkinson's disease using time–frequency analysis and deep learning. *Biomedical Signal Processing and Control*, 78, 103883.
- [9] Altan, G., Yayık, A., & Kutlu, Y. (2021). Deep learning with ConvNet predicts imagery tasks through EEG. *Neural Processing Letters*, 53(4), 2917-2932.
- [10] da Silva Lourenço, C., Tjepkema-Cloostermans, M. C., & van Putten, M. J. (2021). Efficient use of clinical EEG data for deep learning in epilepsy. *Clinical neurophysiology*, 132(6), 1234-1240.
- [11] Bao, Y., & Li, H. (2021). Machine learning paradigm for structural health monitoring. *Structural Health Monitoring*, 20(4), 1353-1372.
- [12] Tan, X., Sun, X., Chen, W., Du, B., Ye, J., & Sun, L. (2021). Investigation on the data augmentation using machine learning algorithms in structural health monitoring information. *Structural Health Monitoring*, 20(4), 2054-2068.
- [13] Malekloo, A., Ozer, E., AlHamaydeh, M., & Girolami, M. (2022). Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights. *Structural Health Monitoring*, 21(4), 1906-1955.
- [14] Muin, S., & Mosalam, K. M. (2021). Structural health monitoring using machine learning and cumulative absolute velocity features. *Applied Sciences*, 11(12), 5727.
- [15] Entezami, A., Shariatmadar, H., & De Michele, C. (2022). Non-parametric empirical machine learning for short-term and long-term structural health monitoring. *Structural Health Monitoring*, 14759217211069842.
- [16] Tan, X., Chen, W., Zou, T., Yang, J., & Du, B. (2022). Real-time prediction of mechanical behaviors of underwater shield tunnel structure using machine learning method based on structural health monitoring data. *Journal of Rock Mechanics and Geotechnical Engineering*.
- [17] Cotorogea, B. P. F., Marino, G., & Vogl, S. (2022). Data driven health monitoring of Peltier modules using machine-learning-methods. *SLAS technology*.
- [18] Sarmadi, H., & Yuen, K. V. (2022). Structural health monitoring by a novel probabilistic machine learning method based on extreme value theory and mixture quantile modeling. *Mechanical Systems and Signal Processing*, 173, 109049.
- [19] Dipietrangelo, F., Nicassio, F., & Scarselli, G. (2023). Structural Health Monitoring for impact localisation via machine learning. *Mechanical Systems and Signal Processing*, 183, 109621.
- [20] Santarsiero, G., Mishra, M., Singh, M. K., & Masi, A. (2021). Structural health monitoring of exterior beam–column subassemblies through detailed numerical modelling and using various machine learning techniques. *Machine Learning with Applications*, 6, 100190.