Forecasting hyponatremia in hospitalized patients using multilayer perceptron and multivariate linear regression techniques

Prasannavenkatesan Theerthagiri

Department of Computer Science and Engineering, GITAM School of Technology, GITAM University, Bengaluru, India

Correspondence
Prasannavenkatesan Theerthagiri, Department of Computer Science and Engineering, School of Technology, GITAM University, Bengaluru 561203, Karnataka, India.
Email: vprasann@gitam.edu

Summary
The percentage of patients hospitalized due to hyponatremia is getting higher. Hyponatremia is the deficiency of sodium electrolyte in the human serum. This deficiency might indulge adverse effects and also be associated with longer hospital stay or mortality if it was not actively treated and managed. This work predicts the futuristic sodium levels of patients based on their history of health problems using a multilayer perceptron (MLP) and multivariate linear regression (MLR) algorithm. This work analyzes the patient’s age, information about other diseases such as diabetes, pneumonia, liver disease, malignancy, pulmonary, sepsis, SIADH, and a sodium level of the patient during admission to the hospital. The results of the proposed MLP algorithm is compared with the MLR algorithm-based results. The MLP prediction results generate 23%–72% of higher prediction results than the MLR algorithm. Thus, the proposed MLP algorithm has produced 57.1% of the reduced mean squared error rate than the MLR results on predicting future sodium ranges of patients. Further, the proposed MLP algorithm produces 27%–50% of the higher prediction precision rate. Therefore, the MLP algorithm can be used for forecasting patient’s hyponatremia.

KEYWORDS
arginine vasopressin, hyponatremia, MLP, prediction, sodium electrolyte

1 INTRODUCTION
Once the human population is suffered from life-threatening diseases such as stroke, central nervous diseases, chronic diseases, renal electrolytes imbalance, then the possibility of occurrence of the same or indulging side effects is higher. The poor diagnosis, treatments, or other health problems of the patients’ might increases the chances of occurrence of the same for the second time. These kinds of deadly diseases should be actively overserved and managed with intensive care.

The homeostatic levels of the human body should be maintained within the normal range.1 The healthy adult human body should have the following range; the serum sodium (Na+) as 136–145 mmol/L; serum potassium (K+) as 3.3–4.5 mmol/L; serum chloride (Cl–) as 96–108 mmol/L; urine sodium as 40–220 mmol/day; urine potassium as 25–125 mmol/day; urine serum chloride as 110–250 mmol/day.2,3 Any abnormalities in these recommended ranges might indulge adverse effects in the human body, and its severity is based on the level of homeostatic concentration or dilution.

The inappropriate antidiuretic hormone secretion might leads to cause the rapid loss of sodium in the serum of humans. The loss of serum sodium concentration in human blood is called as hyponatremia. The sodium level in the blood is diluted by the excess water intake and fre-
quent urination. Typically, hyponatremia is a common disorder and is being observed in many hospitalized patients. It can lead to mortality of the patient if the required essential treatments are not given. Even with adequate intensive treatments, the percentage of mortality is high for patients with hyponatremia. Hyponatremia is also found in patients with the excessive Arginine Vasopressin (AVP) hormone in the plasma of the blood.

Hyponatremia can be categorized into acute and chronic based on the concentration rate of the serum sodium. The too lower serum sodium concentration is very deleterious, and it can affect the normal functions of other organs of the human body. Typically, the most frequent hyponatremia might leads to serious complications in the central nervous system. Patients with acute hyponatremia should be given with intensive and prompt treatments; otherwise, it might lead to life-threatening. Chronic hyponatremia develops the non-renal diseases in other organs of the human body, which often increases morbidity and mortality.

The poor prognosis and treatments are given to the patients with hyponatremia lead to increase the length of hospital stay, especially for heart patients. Hyponatremia might be managed optimally with prompt treatments to the patients, which may increase the serum sodium concentration and reduces the severity and in-hospital mortality. The prompt diagnosis and optimal recognized treatments of hyponatremia to the patients might help to improve their body conditions to normal and also allows them to the normal hospitalization, reduces the duration of hospitalization, and hospitalization costs as well.

Increasing serum sodium concentration too rapidly might lead to life-threatening diseases and sequel side effects. The serum sodium must be increased by the recommended level only; otherwise, it increases the possibility of fatal heart or neurologic disorders. One of the major therapies provided for hyponatremia is associated with blocking the actions of the AVP receptors. It tries to resist the elimination of electrolytes during the urination; it might raise the serum sodium concentration.

The serum sodium concentration should be managed within the recommended range, also based on the homeostatic mechanisms, AVP, renal water excretion, and thirst. Any abnormal changes with the electrolytes of blood, renal chemical levels, and water level could lead to hyponatremia. Even a small unbalance in the renal functions, which was not treated carefully could increase the possibility of consistent morbidity or on the severe condition, maybe death as well. The adverse effects of hyponatremia on the human population motivate the research work on the prediction of future sodium ranges and the possibility of readmission due to hyponatremia on patients. This research work uses the regression and neural networks for the analysis of sodium ranges.

The remainder of the section is organized as follows. Section 2 reports the existing literature works about hyponatremia—the symptoms, causes, existing diagnosis, treatments, effects of hyponatremia. Section 3 explains the proposed methodology and its stepwise details to predict the future sodium ranges for the patients under the hyponatremia treatment. The prediction results, analysis, and evaluation of the proposed methodology are presented in Section 4. Finally, Section 5 concludes with future enhancements.

2 LITERATURE SURVEY

This section summarizes a brief review of existing literacy works that address the importance of hyponatremia and appropriate treatments. Numerically the hyponatremia is defined as the incident that the serum sodium concentration (Na+) reduces below to 136 (normal range 136–145) mEq/L. This type of disorder is common in many hospitalized patients. As mentioned earlier, excessive water retention in the body or persistent urination causes the disorder of dilutional hyponatremia. The water intake should not exceed the kidneys’ excretory capacity. If exceeds, it causes the sodium dilution in the serum, which may lead to hyponatremia, hypo-tonicity, and hypo-osmolality.

Typically, the patients will drink more water if they are affected by the syndrome of inappropriate antidiuretic hormone secretion (SIADH); since the SIADH patients always feel thirst. The treatments and medications are provided to the patients with hyponatremia based on their age, severity, hormone conditions, renal dysfunction, chronic disease, adrenal level, and nervous system. A variety of drugs associated with the treatment of hyponatremia are sulfur-containing diuretics drugs, vasopressin receptor antagonist drugs, tolvaptan, desmopressin, and so forth; these drugs might reduce the risks related to the hyponatremia. The false prognosis of low serum sodium concentration is termed as pseudo-hyponatremia.

It can happen when the patients have extreme hyperlipidemia or hyperproteinemia, and it is identified by flame photometry or indirect potentiometry.

In most of the cases, the excessive AVP secretion with the absence of elevated plasma osmolality is the major issue that causes the hyponatremia. The kidney may retain the water because of the elevated AVP secretion. Thus, the reduced water excretion and increased AVP concentration have a direct association with them. The patients having too-much AVP secretion and exceeding their water intake by 800 mL/day might cause water retention and also it dilutes the fluid compartments, which leads to cause hyponatremia.

Based on age, the occurrence of hyponatremia increases. The documented studies suggest that the old aged people have a higher percentage, such that the 60-aged population has 53% of this disorder in a year. Further, SIADH disorder is most common in the majority of hospitalized patients with hyponatremia. Valproic acid plays a significant role in controlling Na+ channels. It is an 8-carbon 2-chain fatty acid, and the human liver metabolizes it. The valproic acid mediates for the management and recovery of serum sodium concentration through the depolarization of the spinal cord and cortical neurons for the patients.
Typically, its therapeutic range should not exceed 50–100 mcg/mL per dosage. In contrast, its range produces positive responses with minimal side effects experienced by the patients. Also, the level of valproic acid in the blood level should be monitored and altered within the recommended range based on the patients’ responses in a less frequent but regular interval. The time taken to effect the valproic acid on the human body varies between the patients.

The subtherapeutic/supratherapeutic levels of valproic acid to patients might make them to the risk conditions or indulge toxic side effects. If required, the additional testing/treatments should be given to the patients based on medication’s effectiveness/ineffectiveness, patient’s side effects, effects on the central nervous system, and other complications faced by the patients. Further, the patients with multiple medications should be taken additional care with continuous evaluation of valproic acid level because of the interaction of other medications. Also, the protein level in the blood also essentially be monitored because the supratherapeutic valproic acid level has more influence on it.

The research gap in predicting future sodium range of the patients motivates this research work and analysis of the possibility of readmission due to hyponatremia. Table 1 summarizes the top direct evident chemicals and genes which are associated with hyponatremia. The Interferon-gamma (IFNG), AVP, Oxytocin (OXT), TRPV4 (Transient Receptor Potential Cation Channel Subfamily V Member 4) are most evident genes and hormones with the hyponatremia.

| S. No. | Chemical                        | Inferring Genes/Hormones |
|-------|--------------------------------|--------------------------|
| 1.    | Valproic acid                   | AVP, IFNG, OXT           |
| 2.    | Cisplatin                       | IFNG                     |
| 3.    | Dinoprost                       | IFNG, OXT                |
| 4.    | Clonidine                       | AVP, IFNG                |
| 5.    | N-Methyl-3,4-methylenedioxyamphetamine | TRPV4               |
| 6.    | Chlorpropamide                  | AVP                      |
| 7.    | Fluoxetine                      | AVP, IFNG, OXT           |
| 8.    | Colchicine                      | AVP, IFNG                |
| 9.    | Mesna                           | AVP                      |
| 10.   | Fluvoxamine                     | AVP                      |
| 11.   | Indomethacin                    | AVP, IFNG, TRPV4         |
| 12.   | Hydrochlorothiazide             | AVP                      |
| 13.   | Enalapril                       | AVP                      |
| 14.   | Haloperidol                     | AVP                      |
| 15.   | Ifosfamide                      | AVP                      |
| 16.   | Carbamazepine                   | IFNG                     |
| 17.   | Amiloride                       | IFNG                     |
| 18.   | Isoproterenol                   | AVP, IFNG                |
| 19.   | Spironolactone                  | IFNG                     |
| 20.   | Fluourouracil                   | IFNG                     |
| 21.   | Vincristine                     | AVP, IFNG, TRPV4         |
| 22.   | gemcitabine                     | IFNG                     |
| 23.   | Cyclophosphamide                | IFNG                     |
| 24.   | Omeprazole                      | IFNG                     |
| 25.   | Ibuprofen                       | IFNG                     |
| 26.   | Tacrolimus                      | IFNG                     |
| 27.   | Theophylline                    | IFNG                     |
This section describes the methodologies for predicting the future sodium ranges of the patients affected by the hyponatremia. The multilayer perceptron (MLP) and multivariate linear regression (MLR) techniques are adopted in this work for the prediction of future values of the sodium.

The proposed methodology analyzes and predicts the futuristic sodium levels of patients based on their history of health conditions. This research work also facilitates the way to forecast the possibility of the occurrence of hyponatremia once again using artificial intelligence-based algorithms. Further, to emulate the futuristic healthcare of the patients based on their previous history of illness/diseases, the proposed research work had developed an MLR and MLP-based future health prediction algorithm.

### 3.1 Dataset

The dataset of this research is obtained from the Cerner Health Facts database. The dataset is about hospitalized patients during January 2000 and November 2014 collected from various clinics and hospitals in the United States. The dataset contains 1,048,576 numbers of patients’ information. The dataset contains features of patients such as patient’s length of stay, hypertension, coronary_artery_disease, heart_failure, chronic_kidney_disease, end_stage_renal_disease, cirrhosis, chronic_liver_disease, copd, lung_cancer, adrenal_insufficiency, hypothyroidism, depression, dementia, myocardial infarction, peripheral_vascular_disease, cerebrovascular_disease, rheumatologic_disease, peptic ulcer disease, metastatic_cancer_disease, diabetes_complication, hemiplegia, glucose_level during 24 h before admission, glucose_levels during 24 h before admission excluding values ≤20 and >2000, sodium_levels during 24 h before admission, race, gender, sodium_levels adjusted for glucose, serum_sodium_categories, pneumonia, malignancy, pulmonary, sepsis, SIADH, and outcomes.9

The details of the patients such as age (A), gender (G), information about diabetes (D), pneumonia (P), liver_disease (L), malignancy (M), pulmonary (Pu), sepsis (Se), SIADH (S), and sodium level (Na) during admission are taken for this research analysis. The patients are grouped into four categories based on their ages. The age groups are (i) 18 to <45, (ii) 45 to <65, (iii) 65 to <75, (iv) ≥75 years old; the sodium range is categorized as 1 for <120; 2 for ≥120 to <125; 3 for ≥125 to <130; 4 for ≥130 to <135; 5 for ≥135 to <138; and 6 for ≥138 to <140. Similarly, based on gender, the patients are grouped, namely male and female, for the training and learning by the MLR and MLP algorithms.

The data preprocessing and cleaning process removes the missing and outliers data values from the dataset. The resulted dataset after preprocessing is reduced to one million patient records with the above listed 10 required relevant features of patient details. In this dataset, approximately about 49,658 records are missing the essential required patient details. The numerical features from the dataset are taken as the input attributes (gender, age, diabetes, pneumonia, liver_disease, malignancy, pulmonary, sepsis, SIADH), and one feature (Sodium range) is considered as the output attribute. A sample of the patient’s information is presented in Table 2.

### 3.2 MLR algorithm

The MLR is the knowledge-based learning algorithm, works based on the training of the dataset. The MLR can be defined as an improved knowledge-driven expert system: it generates a linear hypothesis and determines the weights for the given variable based on the learning process.
and parameters. The MLR algorithm is a statistics-based model. The MLR algorithm describes the relationship between two or more dependent and independent variables based on the given dataset.

The MLR algorithm computes and analyzes the regression for producing the optimized or appropriate results. It also determines the correlation and assesses the testing/validation and usefulness of the model using various stages of the MLR model. The MLR can be defined as denoted in Equation (1).

\[
\text{MLR}(a, b, k) = \phi + \alpha_1 A + \alpha_2 G + \alpha_3 D + \alpha_4 P + \alpha_5 L + \alpha_6 M + \alpha_7 Pu + \alpha_8 Se + \alpha_9 S + \alpha_{10} Na
\]

In Equation (1), \(\alpha_1, \ldots, \alpha_{10}\) are the set of additive predictor functions, \(\phi\) is the intercept associated with each function, the variable ‘A’ represents the patient’s age, The gender of the patient is denoted as ‘G’, the variable ‘D’ denotes the patients with diabetes, patients with pneumonia is given as ‘P’, ‘L’ gives the patients with liver disease, ‘M’ is malignancy, ‘Pu’ is pulmonary, ‘Se’ is sepsis, ‘S’ is SIADH, ‘Na’ is the sodium level of the patient during admission to the hospital.

### 3.3 Multilayer perceptron (MLP)

This work proposes and develops the MLP-based future sodium prediction algorithm. The MLP algorithm works using the artificial neural network. Typically, in a simple three-layer network, there will be an input layer in the first, a hidden layer in the middle, and the output layer in the last. In the input layers, the input dataset will be fed. The output layer contains the output, which is generated based on the dataset given to the input layer and computations from the hidden layer. The number of the hidden layer can be varied (increased/decreased) based on the complexity of the given problem.

In general, the main objective of any neural network model is to optimize or approximate the given function of the proposed MLP’s classification technique. When the loss function has a high value, the MLP does not make an accurate classification or prediction solution to the given problem, and otherwise, it is vice versa.

In the proposed MLP algorithm, the output of the previous layer is the input to the next layer. Such that the layers of the MLP of fully connected with each other layers in the network. Thus, each unit functions of the layer are always connected to all other layers’ unit function in the neural network. Each layer’s unit functions (i.e., weights and other sets of parameters) are independent of other layer’s unit functions. It means the weights of each layer’s unit functions are unique. Further, the MLP network defines the loss function, which can measure the performance (sodium prediction) of the proposed MLP’s classification technique. When the loss function has a high value, the MLP does not make an accurate classification or prediction solution to the given problem, and otherwise, it is vice versa.

Figure 1 depicts the flow of the proposed MLP-based future sodium prediction algorithm. Firstly, one million patients hyponatremia dataset collected from the hospitals. The outlier and missing data are removed from the hyponatremia dataset using the imputation process. Then the 0.5
million hyponatremia dataset is given to the input layer as the input, and it is trained by using the MLP algorithm. To obtain a better optimal prediction model, the number of hidden layers is varied from two to twenty by the unit of two. This MLP learning and training process gives the prediction results in the output layer.

The prediction results are evaluated for the quality check of the prediction by using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Again, the error performance metrics such as mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean absolute relative error (MARE), and root mean squared relative error (RMSRE) are computed for the prediction results. The AIC, BIC analysis, and error performance metrics computation resulted as the best quality future sodium prediction dataset.34,35 The performance results of the future sodium prediction dataset are analyzed with the precision rate and compared with other existing results.

4  RESULTS EVALUATION AND DISCUSSION

This section gives detailed results analysis, evaluation of performance metrics, and comparative result analysis. The Anaconda Jupiter notebook and various libraries of scikit-learn has been used for the implementation of the proposed work with i3 processor; 3 GB RAM system. The taken dataset has been split into 70%, 15%, and 15% for the training, validation, and testing respectively. To evaluate and validate the performance of the machine learning model, resampling methods are adopted. This method estimates the prediction ability on the machine learning algorithm on new unseen input data. In this work, the 'k' value is chosen as 10; therefore, it can be called as a 10-fold cross-validation resampling method. The 10-fold cross-validation method intends to reduce the bias of the prediction model.

4.1 Results of MLR algorithm

In order to predict the future sodium range, the MLR algorithm trains the 0.5 million datasets containing the patients’ information such as age (A), gender (G), information about diabetes (D), pneumonia (P), liver-disease (L), malignancy (M), pulmonary (Pu), sepsis (Se), SIADH (S), and sodium level of the patients during admission to the hospital (Na). The MLR algorithm trains the 0.5 million patients dataset and performs the testing and validation operations on the given dataset and finally produces the dataset for the future sodium range values.

In order to determine the quality dataset for the prediction of the future sodium values, generalized regression parameters such as (a, b, k) are varied; where the parameter a is the number of input channels, b is the number of output channels, and k represents the delay value. The (a, b, k) parameters are varied as (1,2,1), (1,2,2), (2,2,2), (2,2,3), (2,3,2), (2,3,3), (3,2,1), (3,2,2), and (3,3,2). Table 2 summarizes the results of error performance metrics such as MSE, RMSE, MAE, MARE, and MSRE for the MLR parameters (a, b, k) using the MLR algorithm.34

As per the definition of MSE, the resultant lowest MSE value among the different (a, b, k) parameter gives a feasible, realistic solution. Therefore in Table 3, the lowest MSE value that is (a, b, k) parameter (2,2,1) are highlighted with boldface. The error performance metric values for the MSE, RMSE, MAE, MARE, and MSRE for the MLR parameters (2,2,1) are 0.1894, 0.4352, 0.4162, 0.1036, and 0.1121 respectively; it is the lowest among other (a, b, k) parameters. Therefore, the corresponding dataset of the (2,2,1) parameter is considered as the appropriate and optimistic solution for the given hyponatremia patient dataset.

| Error metrics/(a, b, k) | MSE      | RMSE    | MAE   | MARE  | RMSRE  |
|------------------------|----------|---------|-------|-------|--------|
| (1,2,1)                | 0.215    | 0.4637  | 0.4386| 0.107 | 0.1148 |
| (1,2,2)                | 0.2643   | 0.5141  | 0.4849| 0.1173| 0.1247 |
| (2,2,1)                | 0.1894   | 0.4352  | 0.4162| 0.1036| 0.1121 |
| (2,2,2)                | 0.2668   | 0.5165  | 0.4908| 0.1192| 0.1263 |
| (2,2,3)                | 0.2923   | 0.5406  | 0.5097| 0.123 | 0.1303 |
| (2,3,2)                | 0.2585   | 0.5084  | 0.4794| 0.1161| 0.1236 |
| (2,3,3)                | 0.5028   | 0.7091  | 0.6688| 0.1611| 0.1699 |
| (3,2,1)                | 0.2152   | 0.4639  | 0.439 | 0.1071| 0.1149 |
| (3,2,2)                | 0.2671   | 0.5168  | 0.491 | 0.1193| 0.1264 |
| (3,3,2)                | 0.2585   | 0.5084  | 0.4795| 0.1161| 0.1236 |

**TABLE 3** Performance error metrics resulted by MLR
### TABLE 4  Results of AIC and BIC metrics for MLR

| Criterion/\(a, b, k\) | AIC      | BIC      |
|-----------------------|----------|----------|
| (1,2,1)               | 731,809.6| 731,776.2|
| (1,2,2)               | 811,329.74| 813,626.37|
| (2,2,1)               | 591,230.55| 593,907.18|
| (2,2,2)               | 611,590.78| 662,324.148|
| (2,2,3)               | 672,218 | 672,351.3 |
| (2,3,2)               | 600,211 | 600,215 |
| (2,3,3)               | 768,794.9| 768,828.3|
| (3,2,1)               | 733,165.4| 733,132.1|
| (3,2,2)               | 720,541.85| 780,735.22|
| (3,3,2)               | 822,211.59| 901,378.22|

Besides, the results of AIC and BIC metrics for the MLR algorithm given in Table 4 confirm that the \(2,2,1\) parameter produces the lowest metric results. As per the AIC and BIC definition, the better quality and stable can be given by the lowest AIC, or BIC valued corresponding dataset. In Table 4, the boldfaced \(2,2,1\) parameter gives the lowest AIC and BIC metrics, such as 2211.489 and 2214.856 respectively. The resultant values of AIC, BIC, and error metrics confirm that the dataset of \(2,2,1\) parameter produces better prediction results. Therefore, the \(2,2,1\)’s corresponding dataset is taken as the result of the future sodium prediction values.

### TABLE 5  Performance error metrics resulted by MLP

| Metrics/neurons | MSE   | RMSE  | MAE   | MARE  | RMSRE  |
|-----------------|-------|-------|-------|-------|--------|
| 2               | 0.052 | 0.2281| 0.1521| 0.0418| 0.073  |
| 4               | 0.1012| 0.3181| 0.2791| 0.071 | 0.0919 |
| 6               | 0.0227| 0.1505| 0.0691| 0.0196| 0.0441 |
| 8               | 0.0942| 0.3069| 0.273 | 0.07  | 0.0898 |
| 10              | 0.0563| 0.2372| 0.1693| 0.0492| 0.1017 |
| 12              | 0.1092| 0.3304| 0.2975| 0.0752| 0.0956 |
| 14              | 0.1014| 0.3184| 0.2832| 0.0728| 0.0967 |
| 16              | 0.1018| 0.3191| 0.2862| 0.0727| 0.0914 |
| 18              | 0.0806| 0.2839| 0.248 | 0.0636| 0.0773 |
| 20              | 0.1005| 0.317 | 0.2848| 0.0722| 0.0894 |

Besides, the results of AIC and BIC metrics for the MLP algorithm given in Table 4 confirm that the \(2,2,1\) parameter produces the lowest metric results. As per the AIC and BIC definition, the better quality and stable can be given by the lowest AIC, or BIC valued corresponding dataset. In Table 4, the boldfaced \(2,2,1\) parameter gives the lowest AIC and BIC metrics, such as 2211.489 and 2214.856 respectively. The resultant values of AIC, BIC, and error metrics confirm that the dataset of \(2,2,1\) parameter produces better prediction results. Therefore, the \(2,2,1\)’s corresponding dataset is taken as the result of the future sodium prediction values.

### 4.2  Results of MLP algorithm

The MLP algorithm trains the dataset containing the 0.5 million patients’ information such as age \(A\), gender \(G\), information about diabetes \(D\), pneumonia \(P\), liver-disease \(L\), malignancy \(M\), pulmonary \(Pu\), sepsis \(Se\), SIADH \(S\), and sodium level \(Na\) of the patients during admission to the hospital. In order to determine the quality dataset for the prediction of the future sodium values, the number of hidden neurons is varied from two to twenty.\(^{36,37}\) Table 5 summarizes the resultant performance error metrics values such as MSE, RMSE, MAE, MARE, and MSRE for the MLR algorithm with several different neurons.

In Table 5, the lowest MSE value is highlighted as neuron 6. The error performance metric values for the MSE, RMSE, MAE, MARE, and MSRE for the neuron 6 by MLP algorithm are 0.0227, 0.1505, 0.0691, 0.0196, and 0.0441 respectively; it is the lowest among other neurons. Therefore, the corresponding dataset of the neuron 6 is considered as the appropriate and optimistic solution for the given hyponatremia patient dataset.

Moreover, the results of AIC and BIC metrics for the MLP algorithm is summarized in Table 6. It confirms that the neuron 6 generates the lowest metric results. In Table 3, the boldfaced \(2,2,1\) parameter gives the lowest AIC and BIC metrics, such as 525,639.24 and 525,605.88 respectively. The resultant values of AIC, BIC, and error metrics confirm that the dataset of neuron 6 produces better prediction results. Therefore, neuron 6’s corresponding dataset is considered as the result of the future sodium prediction values in this scenario.
### Table 6 Results of AIC and BIC Metrics for MLP

| Criterion/neurons | AIC          | BIC          |
|-------------------|--------------|--------------|
| 2                 | 950,044.093  | 950,021.848  |
| 4                 | 759,348.725  | 759,326.48   |
| 6                 | 525,639.24   | 525,605.88   |
| 8                 | 780,974.062  | 780,951.818  |
| 10                | 950,819.348  | 950,797.103  |
| 12                | 708,582.641  | 708,560.397  |
| 14                | 777,258.361  | 777,236.116  |
| 16                | 764,183.817  | 764,161.572  |
| 18                | 832,843.723  | 832,821.478  |
| 20                | 757,474.068  | 757,451.823  |

### Figure 2 Comparison of Na+ prediction results by MLP and MLR (Critical)

#### 4.3 Result analysis for MLR and MLP algorithm

The future sodium prediction results obtained by MLR and MLP are compared for the analysis of accurate prediction of results. Figure 2 gives the comparative future sodium prediction results by using the techniques MLP and MLR for the critical hyponatremia patients. The observed and predicted results of the patient’s serum sodium range, such as less than 120, 120 to 125, and 126 to 130, are depicted in Figure 2. For the sodium range less than 120 categories, the total number of observed patients are 2568; the proposed MLP algorithm had predicted the total number of patients under less than 120 category as 2537; whereas the MLR had predicted it as 5399 patients for the less than 120 categories.

The proposed MLP algorithm had produced higher accuracy of prediction, in which the prediction difference with the observed results is 1.21% only. In contrast, the MLR algorithm has a prediction difference, with the observed results as 71%. The total number of observed patients for the sodium range 120 to 125 category is 9639; the proposed MLP algorithm had predicted the total number of patients for the 120 to 125 category as 10,554; whereas the MLR had predicted it as 15,370 patients for the same category. In this case, the prediction difference with the observed results MLP and MLR algorithms are 9% and 45% respectively.

Similarly, for the sodium range 126–130 category, the total number of observed patients are 45,024; the proposed MLP algorithm had predicted the total number of patients for the same sodium category as 42,797; whereas, the MLR had predicted it as 25,711 patients for the same category. Correspondingly, the MLP algorithm has a lower prediction difference with the observed results as 5%; whereas, the MLR algorithm has a higher prediction difference with the observed results as 54.6%.

For the stable hyponatremia patients, the comparative future sodium prediction results by using the techniques MLP and MLR are depicted in Figure 3. The total number of observed patients for the sodium range 131–135 category is 300,187; the proposed MLP algorithm had predicted the total number of patients for the same category as 305,875; whereas, the MLR had predicted it as 385,388 patients for this category. In this case, the prediction difference with the observed results MLP and MLR algorithms are 1.8% and 24.8% respectively.

Similarly, for the sodium range 136–138 category, the total number of observed patients are 142,582; the proposed MLP algorithm had predicted the total number of patients for the same sodium category as 13,827; whereas, the MLR had predicted it as 68,117 patients for the same
Comparison of Na+ prediction results by MLP and MLR (Stable)

![Comparison of Na+ prediction results by MLP and MLR (Stable)](image)

**TABLE 7** Error metrics result analysis for MLR and MLP

| Metrics/algorithm | MSE   | RMSE  | MAE   | MARE  | RMSRE |
|-------------------|-------|-------|-------|-------|-------|
| MLR               | 0.1894| 0.4352| 0.4162| 0.1036| 0.1121|
| MLP               | 0.0227| 0.1505| 0.0691| 0.0196| 0.0441|

Age wise hyponatremia prediction by MLP

![Age wise hyponatremia prediction by MLP](image)

category. Similarly, the MLP algorithm has a lower prediction difference with the observed results as 3.1%; whereas, the MLR algorithm has a higher prediction difference with the observed results as 70.6%.

Table 7 gives the best results of the performance error metric for the proposed MLP and MLR algorithms. The appropriate stable dataset (2,2,1) discovered by the MLR algorithm has the performance error metrics such as MSE, RMSE, MAE, MARE, and MSRE as 0.1894, 0.4352, 0.4162, 0.1036, and 0.1121 respectively. Similarly, the quality dataset (neuron 6) determined by the proposed MLP algorithm has the performance error metrics such as MSE, RMSE, MAE, MARE, and MSRE as 0.0227, 0.1505, 0.0691, 0.0196, and 0.0441 respectively. Henceforth, these confirm that the proposed MLP algorithm had generated the lower error rates than the MLR error rates.

Figure 4 gives the patients’ age wise hyponatremia prediction results based on the MLP algorithm. From this pie chart, it can be seen that the patients with below the age of 45 are getting affected by the hyponatremia is significantly less (i.e.,) 17%. When the patients’ age is above 75 or the patients’ age is 46–65, then the number of patients getting affected by the hyponatremia is very high (i.e.,) 32%. Between the age of 66–75, there are 19% of patients had treatment for hyponatremia. Figure 5 illustrates the pie chart analysis on MLP-based hyponatremia prediction results based on gender. From Figure 5, it is clear that female patients are highly suffered from hyponatremia than male patients. Such that 54% of female patients had treatments for hyponatremia, whereas 46% of male patients only had treatments for hyponatremia from the dataset of 0.5 million patients.

The prediction results of the proposed MLP algorithm suggest that it predicts the future sodium level of the patients based on their disease history. Therefore, the anticipated benefits of the proposed MLP algorithm will be useful to patients with a similar disease history. Such that the
4.4 Computation and analysis of the precision rate

In order to analyze the accuracy of the prediction results of the MLR and MLP algorithms, the precision rate is calculated. The prediction precision rate is the percentage of the root of the squared difference rate between the predicted and observed results calculated using the Euclidean distance. It is given in Equations (2), (3) as,

\[ PR = \left( 1 - \frac{EUC_{\text{Dis}}}{EUC_{\text{max}}} \right) \times 100 \]  

(2)

\[ EUC_{\text{Dis}} = \sqrt{\text{Predicted} - \text{observed}}^2 \]  

(3)

where, EUC_{\text{Dis}} denotes the Euclidean distance and EUC_{\text{max}} represents the maximum Euclidean distance among predicted and observed serum sodium results. Table 8 gives the accuracy of the prediction results using precision rate analysis. In Table 8, for the different serum sodium range such as less than 120, 120–125, 126–130, 131–135, and 136–138, the proposed MLP algorithms have the prediction precision rates as 98.7928, 90.5073, 95.0537, 98.1052, and 96.9456. In contrast, the MLP algorithms have prediction precision rates as 51.2414, 40.5436, 57.1051, 71.6174, and 47.7739. The proposed MLP algorithm has a prediction accuracy of 90.5%–98.7%, whereas the MLR algorithm has a prediction accuracy of 40.5%–71.6%. Such that the proposed MLP algorithm has 27%–50% of higher precision rate on predicting the future sodium range of the patients.

Table 9 summarizes the percentage of difference (PD) among the observed results, MLR algorithm-based prediction results, and MLP algorithm-based prediction results. In this analysis, the proposed MLP algorithm was produced only 1.2%–9% of the difference with the observed results. In contrast, the MLR algorithm-based prediction results have a 24.8%–71% of the difference with the observed results since the proposed MLP algorithm was produced 23.6%–62% of the reduced percentage of difference with the observed results. As well as the MLP predictions has improved results of 23%–72% as compared with MLR prediction results. The confusion matrix for the MLP algorithm for the less than 120 Na+ range is given in Figure 6. The confusion matrix shows the higher prediction for hyponatremia and other sodium range prediction ranges as 0.89 and 0.7.
TABLE 9 Percentage of difference for observed, MLR, and MLP results

| Results/Na+ range | MLP with observed results | MLR with observed results | MLP with MLR results |
|-------------------|--------------------------|--------------------------|----------------------|
| <120              | 1.2145                   | 71.0682                  | 72.127               |
| 120–125           | 9.0625                   | 45.8315                  | 37.1548              |
| 126–130           | 5.0717                   | 54.6066                  | 49.8803              |
| 131–135           | 1.877                    | 24.8553                  | 23.0051              |
| 136–138           | 3.1018                   | 70.6838                  | 67.9545              |

FIGURE 6 Confusion matrix of hyponatremia prediction by MLP

4.5 | Data sharing and data accessibility

The data that support the findings of this study are openly available in "DRYAD" at https://datadryad.org/stash/dataset/doi:10.5061/dryad.89nn8, reference number.

5 | CONCLUSION AND FUTURE WORK

This work was concentrated on the prediction of future sodium range for the patients based on various health history factors such as age, gender, health problems, and so forth, in order to predict the hypo/hypernatremia. The proposed MLP algorithm has produced an accurate future serum sodium prediction range than the MLR algorithm. The MLR algorithm has a prediction accuracy rate of 41%–72%, whereas the MLP neural network algorithm has an accurate prediction of 91%–99%. The MLP algorithm-based prediction results have 27%–50% of improved prediction accuracy than the MLR algorithm-based prediction results. Moreover, the proposed MLP algorithm-based prediction results in 57.1% of the reduced MSE error rate than the MLR results in predicting future sodium ranges of patients. The outcome of the proposed MLP algorithm-based future health prediction algorithm could be more helpful for physicians and patients to make further decisions based on their health conditions. However, the limitation of this work is it does not focused on the survival rate of the patients after suffering the hyponatremia. The future work will concentrate on forecasting the possibility of re-occurrence of hyponatremia for the patients and survival rate of the patients after hyponatremia with various health parameters using machine learning prediction algorithms.

ORCID
Prasannavenkatesan Theerthagiri https://orcid.org/0000-0003-3420-598X

REFERENCES
1. Castillo JJ, Glezerman IG, Boklage SH, et al. The occurrence of Hyponatremia and its importance as a prognostic factor in a cross-section of cancer patients. BMC Cancer. 2016;16(1):564.
2. Roche diagnostics ISE/Na+, K+, Cl-package insert, 2009-03, V6.
