Uluslararası İleri Doğa Bilimleri ve Mühendislik Araştırmaları Dergisi
Sayı 8, S. 188-196, 9, 2024

Araştırma Makalesi

International Journal of Advanced Natural Sciences and Engineering Researches
Volume 8, pp. 188-196, 9, 2024

Research Article

IJANSER

https://as-proceeding.com/index.php/ijanser
ISSN:2980-0811

Performance Comparison of Learning Models to Predict The In-Hospital Mortality

Mehmet Kurucan*[1]

[1]*Bilgisayar Mühendisliği, Mühendislik Fakültesi, Ardahan Üniversitesi, Türkiye*

**mehmetkurucan@ardahan.edu.tr*

(Received: 01 October 2024, Accepted: 18 October 2024)

(5th International Conference on Innovative Academic Studies ICIAS 2024, 10-11 October 2024)

ATIF/REFERENCE: Kurucan, M. (2024). Performance Comparison of Learning Models to Predict The In-Hospital Mortality, *International Journal of Advanced Natural Sciences and Engineering Researches*, 8(9), 188-196.

Abstract – The prediction probability of in-hospital mortality who are admitted to the Intensive Care Unit (ICU) are calculated in this work by applying two common learning models: artificial neural networks and the hidden Markov model. Both models were applied to the same dataset to ensure a fair comparison. The clinical data supporting these models were carefully selected from the Multi-Parameter Intelligent Monitoring (MIMIC III) database, with a particular emphasis on the ICU domain. Thus, accurately the real-world conditions were reflected. The dataset, comprising 8000 individual records, was divided using cross-validation techniques. Subsequently, the datasets were utilized as training and test sets for each learning model. The effectiveness of the models was evaluated using the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) measure due it aligns well with the fundamental characteristics of the models. Notably, the single hidden layer neural network model produced a ROC value of 0.8927, while the multi-hidden layer model generated a significantly lower ROC value of 0.8691. The hidden Markov model achieved the best result in this study, with a higher ROC value of 0.9038.

Keywords – Intensive Care Unit, Artificial Neural Networks, Hidden Markov Models, Receiver Operating Characteristic Curve, Multi-Parameter Intelligent Monitoring (MIMIC III).

I. INTRODUCTION

Historically, the intensive care unit (ICU) has been the most important aspect of medical facilities. This significance is due to the fact that patients admitted to these units almost always have acute critical conditions, necessitating rigorous and ongoing monitoring of their physiological measurements. These recorded metrics, which include variables such as electrocardiogram traces, blood pressure readings, haematological indices, pharmaceutical treatments, and the resulting pharmacodynamic responses, provide a comprehensive dataset. Using this dataset, reliable prediction of an individual patient's projected mortality risk becomes critical, allowing the medical team to make more precise management decisions.

In this scenario, the use of artificial neural networks (ANNs) and hidden Markov models (HMMs) emerge as the most promising techniques. This class of machine learning algorithms, well-known for their widespread application, is designed to address such predicted quandaries. The prediction of mortality outcomes for ICU patients becomes possible with the use of such networks, offering to increase clinical practitioners' prognostic powers in this vital arena.

In this work, two distinct ANNs and an HMM have been employed to assess the prediction of death outcomes in hospital. These machine learning models give superior performance to predict such outcomes. Authentic datasets were utilized to ensure the verifiability of the results and the accuracy of the models' predictive capabilities. In order to achieve this, relevant data for the study was carefully gathered from a reputable source of freely accessible clinical data called the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC III) Clinical database [1].

In the subsequent analytical process, the dataset were handled with two ML models to add incremental levels of complexity to the study. These models are architecturally distinct. One has hidden layers and the other has hidden states. The impact of these models was compared through performance evaluation using the well-recognized receiver operating characteristic (ROC) curve metric. The ROC curve, which is renowned for its discriminative capability, yields an area under the curve (AUC) value. The values which are approaching 1 indicate superior performance.

Numerous studies have been conducted in relation to the prognostication of hospital mortality through the use of ML models. The use of the ROC curve as a major metric for evaluating model performance has continuously been a complement to this paradigm. For instance, [2] used an ANN-based model in their study, which resulted in an AUC value of 0.743 when viewed in the context of the ROC curve. In a different research, [3] used a random forest technique to their in-hospital mortality forecasting model, resulting in an AUC of 0.862 inside the ROC framework. In contrast, [4] used their suggested model to reach an identical in-hospital mortality estimation, albeit with a minor distinction, with an AUC value of 0.86. [5] broadened the methodological spectrum even further by applying a random forest model to predict death. This project yielded an AUC value of 0.84, capturing its model's discriminatory skills within the ROC demarcation.

Section 2 contains information on the dataset used in this work. Section 3 gives an overview of the methodologies used throughout the study. Section 4 summarises the findings, whereas Section 5 presents the conclusion.

II. MATERIALS AND METHOD

A. DATA MANAGEMENT

In research focused on hospital mortality estimation, the utilization of authentic datasets is critical. Robust training of predictive models proves useful. They offer expert clinicians a critical tool for predicting patient outcomes. Consequently, the datasets used in this study were intently obtained from the freely accessible MIMIC-III repository. Hence, we support the integrity and accuracy of the model training process with this choice.

8000 entries were collected from the MIMIC III database to be used in this study. This comprehensive dataset contains a wide range of information, including demographic characteristics such as age, gender, height, and weight of ICU patients. Additionally, the dataset includes 37 unique time series that monitor the progression of various physiological indicators over time, such as blood pressure, heart rate, and blood glucose levels. We divide the dataset into distinct sections for the training and testing phases. Thus, we did optimize the efficacy of the data-driven models employed. A symmetrical distribution of 4000 records was used for both the training and testing stages. In this way, we promote a fair data allocation during model evaluation.

B. ERRORS IN DATA

During the process of data collection, some of the variables may not be recorded for certain patients, leading to missing values in the dataset. To avoid this issue, all NaN values were accepted as an error and they were replaced with zero value.

C. NOISE AND MISSING DATA

To address the issue of missing data and noise in the dataset, a preprocessing step was applied to the data. The preprocessing step involved removing time series data with more than 50% missing data and imputing missing data with the median value of the corresponding time series data. All missing data were also replaced with zero.

D. POSITIVE AND NEGATIVE CLASSES

The dataset must be separated into positive and negative classes. Because it is necessary to find the ROC values to compare the models in the implementation phase. The ROC curve is produced by plotting the true positive rate against the false positive rate. Thus, the data needs to be separated into positive and negative classes to calculate these values. A criterion is required for this data separation. Our criterion will be whether the patient survived or did not survive during their stay in the ICU. Thus, while dividing the dataset into positive and negative classes, we labelled the data according to this criterion.

E. SEQUENCE PREPARING

For the training of the HMM, observation sequences are required. Therefore, it is necessary to convert the used dataset into observation sequences. This was done by creating time-series data from the records, such as sequences of daily measurements.

F. METHOD

In this study, the application of two different machine learning models is based on the notion that, given the natural characteristics of the dataset used, a simpler model may outperform a more complex model in terms of producing higher predictive accuracy. Therefore, to observe the performance of two well-known machine learning models in predicting in-hospital mortality, we applied the following methods.

G. ANN MODELS

We selected ANN as our first machine learning model due to its widespread use and practicality in class problems. We designed two different models for the ANN. One model contains a single hidden layer, chosen for its simplicity, while the other model includes two hidden layers to increase complexity.

The first model employed the Levenberg-Marquardt training procedure, which is recognized to be the fastest training function. Unfortunately, it necessitates additional memory and computing time. This algorithm is appropriate for solving nonlinear regression issues. The Gradient Descent with Momentum function, on the other hand, was used for the second model to boost the rate of learning. The Mean Square Error performance function was used to optimize both models.

H. ONE HIDDEN LAYER FEED-FORWARD NEURAL NETWORK

ANN used in this model has a single hidden layer with a dynamic allocation of nodes as its architecture. For the purposes of this model, the Levenberg-Marquardt strategy was chosen among the range of training methodologies that are readily available. In parallel, the model's optimisation was controlled by the use of the mean square error criterion. Notably, the output layer was distinguished by the use of the linear transfer function, while the hidden layer was labelled as having the hyperbolic tangent sigmoid function as its activation function.

İ. TWO HIDDEN LAYERS FEED-FORWARD NEURAL NETWORK

The second model used for the purposes of this study uses a configuration with two hidden layers. Unlike the original model, this architecture's training process made use of the Gradient Descent with Momentum optimisation algorithm. The degree of approximation of the model was once more measured using the mean square error. A deviation from the strategy used in the first model is seen in the choice of activation functions. This second model applied the tan-sigmoid transfer function uniformly to both the output layer and the concealed layers.

J. HIDDEN MARKOV MODELS

An HMM is defined as a statistical model where a process runs behind as hidden. It performs the run over particular sequence (or time series data) where the system being modelled is assumed to be a Markov process with unobserved (hidden) states.It is widely used in various fields such as stock market forecasting [6], bioinformatics [7], and natural language processing[8].

HMMs are employed to compute the probability of the given observation data for classification. For this work topic, the binary classification problem can be resulted by training two separate HMMs. Here, we have two HMM models for positive and negative classes. Both models are trained for each class. The workflow is proceeded as follow phases:

- **Training phase**: Both HMMs are trained for each class (positive and negative).
- **Scoring phase**: For each test sequence, calculate the log-likelihood for both the positive and negative HMMs. Here, the decision score is computed by using the log-likelihood ratio.
- **Evaluation phase**: The ROC curve is plotted by using the decision scores. Then, the area under the curve of ROC is calculated to evaluate the performance.

K. NUMBER OF NODES

After defining the procedures guiding the activation and transfer functions in ANN and defining the phases of HMM process, a crucial decision-point that required thought emerged: the choice of the node count. This parameter clearly influences the dynamics of model training and the results that follow. Consequently, a deliberative technique was used to determine the node allocation for ANN by extracting conclusions from a prior study [9]. The node allocation procedure for a particular ANN model was driven by the study's premise, which stipulates the relationship $H = log2$. This mathematical formulation resulted in the first ANN model having 8 nodes and the second ANN model having 13, respectively.

Determining the number of nodes in the HMM can be done with any of two criteria which are the Akaike Information Criterion and the Bayesian Information Criterion. These are statistical approaches employed to balance model fit with complexity. According to these two criteria, 11 nodes suit the HMM model for the problem of this work.

L. CROSS VALIDATION

To evaluate the performance of the ML models, 4000 records were randomly selected from the dataset and split into 5 sub-datasets, each with 800 samples for cross-validation (Awwalu et.al, 2019). The model was trained and tested 5 times, with each sub-dataset used once for testing and the rest for training. The evaluation of the models' performance was subsequently a crucial task. Calculating the AUC of the ROC, a metric naturally appropriate for the assessment of such prediction models, served as the benchmark for this examination.

The computation of a bounded region is essentially required when using the ROC methodology to evaluate the effectiveness of models built on ML. This region corresponds to the space encircled by the curve that the ROC graph has drawn. The range of this area is limited to values between 0.5 and 1, with a value closer to 1 denoting a higher level of predicted accuracy provided by the model. As a result, the model developed for this study, which shows a bias towards a value close to 1, naturally improves the accuracy of its prognostic results.

III. RESULTS

In this section, we represent how to accomplish the objective by trying different parameters during ANNs and HMMs training. Here, our goal is to determine the best parameters to optimise the ROC value. The number of nodes that are used in two ANNs and HMMs are explained in subsection 3.3. Overfitting will occur if a suitable number of nodes used in hidden layers is not specified in the models. To prevent this situation, we aimed for a network size for ANNs that is just large enough and parameters for HMMs to obtain a satisfactory fit without overfitting the training data.

Several training groups are developed to select the optimum network model, as shown in Table 1 and Table 2, with the goal of achieving the suitable number of nodes to maximize the AUC. This method focuses on determining the precise node numbers for ANN and HMM models that offer the highest AUC value.

Table 1. The higher AUC values with the number of optimum number of the nodes for two ANN models

Hidden Layers	**Node Range**	**The Numbers of Node with High AUC**	**AUC Value**
1	8-13	13	0.8927
1	14-19	19	0.8861
1	20-25	22	0.7788
2	8-13	12	0.7451
2	14-19	15	0.8451
2	20-25	24	0.8691
2	26-31	27	0.8157

Table 2: The higher AUC values with the number of optimum number of the nodes for HMM models

Node Range	**The Numbers of Node with High AUC**	**AUC Value**
8-13	11	0.9038
14-19	16	0.8741
20-25	21	0.8357

Although the appropriate number of nodes for the models has been established using the methods described in the previous section, the AUC values of different node values have also been examined in this section to minimize the margin of error. This approach ensures that we avoid overfitting of the models.

The other goal is to calculate the number of iterations for both models (ANNs and HMMs) that should be used during training. The iteration parameter has an important spot to avoid overfitting the model when it is being trained iteratively. When the model overfits, the validation error increases. When the validation error approaches the max_fail value in ANNs and log-likelihood of the validation below a certain threshold in HMMs, the training is terminated. Table 3 and Table 4 present the appropriate max_fail values that yield high AUC values for the one hidden layer ANN and two hidden layers ANN models, respectively. Table 5, on the other hand, shows the suitable iteration value that provides the highest AUC value for the HMM model.

Table 3: The best iteration values of one-hidden layer network

Network and Number of Nodes	**max_fail value**	**AUC Value**
1-8-1	50	0.8075
1-9-1	60	0.8114
1-10-1	50	0.8259
1-11-1	50	0.8347
1-12-1	50	0.8564
1-13-1	50	0.8927
1-14-1	50	0.8763
1-15-1	80	0.8745
1-16-1	80	0.8697
1-17-1	80	0.8637
1-18-1	80	0.8413
1-19-1	80	0.8861

Table 4: The best iteration values of two-hidden layers network

Network and Number of Nodes	max_fail value	AUC Value
1-14-14-1	1200	0.8075
1-15-15-1	1200	0.8451
1-16-16-1	1200	0.8429
1-17-17-1	1200	0.8318
1-18-18-1	1220	0.8317
1-19-19-1	1260	0.8314
1-20-20-1	1200	0.8211
1-21-21-1	1280	0.8196
1-22-22-1	1200	0.8178
1-23-23-1	1200	0.8564
1-24-24-1	1200	0.8691
1-25-25-1	1200	0.7957

Table 5: The best iteration values of HMMs

Number of Nodes	Number of Iteration	AUC Value
8	57	0.8741
9	57	0.8812
10	57	0.8901
11	57	0.9038
12	60	0.8872
13	60	0.8691
14	60	0.8458
15	58	0.8608
16	58	0.8741
17	58	0.8596
18	60	0.8412
19	60	0.8396

In conclusion, to achieve the highest AUC value, we determined the iteration values to prevent overfitting by optimizing both the node values and the training steps. With these two crucial criteria, the parameter values for the optimal model were identified, and higher AUC values were achieved with the HMM on the same dataset during the test results. Naturally, the dataset used and other parameters (such as the number of observation symbols, the use of different transfer and training functions, etc.) will also influence the achieved AUC value. Although HMM is a commonly used ML model for time-series problems, it has been observed that ANN achieves quite similar AUC values on the same task where it is being an alternative ML model. Depending on the purpose of use, complexity of structure, and ease of implementation, these models can be utilized for different objectives.

IV. CONCLUSION

Accurate prognostications are essential given the critical importance of addressing concerns like mortality estimation for hospitalised patients, especially in the context of the intensive care unit (ICU). The search for improved solutions led to an extensive investigation that involved a comparative analysis using the various Machine Learning (ML) models. Two common ML models, which are Artificial Neural Network (ANN) and Hidden Markov Model (HMM), are utilized to this strategy for aiming to improve the accuracy of the predictions by utilising the inherent variability in model configurations.

In order to do this, a comparison analysis of models with various structural features was conducted within the context of the same task. To be more specific, ANN has two diffrerent models which have one and two hidden layers, respectively. Notably, the experimental modifications included different, diverging node numbers, and different training approaches in addition to subtle structural variations. Through these several model iterations, a thorough investigation was conducted to identify the ideal setup that may produce increased precision in mortality estimates, suited to the requirements of the ICU scenario.

For both models that were used, the Receiver Operating Characteristic (ROC) technique was used to assess model performance in the current study. The AUC calculated from the initial model, which only had one hidden layer, produced a result of 0.8927. The AUC determined from the second ANN model, on the other hand, was 0.8691. But on the other hand, a high AUC value, 0.9038, has been achieved with HMM on the same task.

In conclusion, our study shows that ML models have the capacity to predict in-hospital mortality in ICU patients. The performance of the HMM and one-hidden-layer ANN is close to each other. Future research could look into the effect of additional factors or different topologies to increase prediction performance even further.

REFERENCES

[1] Clifford G.D., Scott D.J., Villarroel M. User guide and documentation for MIMIC III database, Version 1.4, Realised date: 02.09.2016. Available: http://www.physionet.org or http://mimic.physionet.org

[2] Hatamabadi, H. R., Abdalvand, A., Abdollahi, F. Shokoohi, H., Bastani, M., The utility of artificial neural network models in predicting the mortality of patients with traumatic brain injury admitted to intensive care unit. Archives of trauma research 2020; 9(1), 30-35

[3] Chen, Y., Li, L., Du, J. Predicting in-hospital mortality of intensive care unit patients using random forest model. Journal of Medical Systems 2019; 43(8), 239.

[4] Shashikumar, SP., Stanley, MD., Sadiq, I., Li, Q., Holder, A., Clifford, GD., Nemati, S. Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. Journal of biomedical informatics 2017; 75, 8-15.

[5] Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med. 2018;24(6):1716 1720.

[6] M. R. Hassan and B. Nath. Stock market forecasting using hidden markov model: a new approach. In 5th international conference on intelligent systems design and applications (ISDA'05), pages 192–196. IEEE, 2005.

[7] V. De Fonzo, F. Aluffi-Pentini, and V. Parisi. Hidden markov models in bioinformatics. Current Bioinformatics, 2(1):49–61, 2007.

[8] V. Nefian, L. Liang, X. Pi, L. Xiaoxiang, C. Mao, and K. Murphy. A coupled hmm for audio-visual speech recognition. In 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 2, pages II–2013. IEEE, 2002.

[9] Mirchandani, G., Cao W. On hidden nodes for neural nets. IEEE Transaction on circuits and systems 1989; 36, 5.