Open Access

Multi-layered deep learning perceptron approach for health risk prediction

Thulasi Bikku[*]

*Correspondence:
thulasi.jntua@gmail.com
Vignan's Nirula Institute
of Technology and Science
for Women, Palakaluru,
Guntur, Andhra Pradesh
522009, India

Abstract

In today's world, due to the increase of medical data there is an interest in data pre-processing, classification and prediction of disease risks. Machine learning and Artificial Intelligence indicates that the predictive analysis becomes part of the medical activities especially in the domain of medical death prevention. The proposed work is focused on supervised learning methods and their capability to find hidden patterns in the real historical medical data. The objective is to predict future risk with a certain probability using Multi-layer perceptron (MLP) method. In the proposed work, MLP based on data classification technique is used for accurate classification and risk analysis of medical data. The proposed method is compared with traditional classification methods and the results show that the proposed method is better than the traditional methods.

Keywords: Multi-layer perceptron, Medial data, Classification, Deep learning, Decision support systems

Introduction

Nowadays, we are often dealing with data that contains many variables, columns, or attributes. Those datasets, which can be huge in terms of variables or rows, are called high-dimensional data. Electronic Health Record (EHR) may have hundreds of input variables, ranging from age, gender, to various lab results. Data collected from Deoxyribonucleic acid (DNA) microarray typically have thousands to ten thousands of features representing gene expression [1]. An image, or a text document, is another example of high-dimensional data in which every pixel or word can be seen as a feature. Processing, analysing, and organizing high-dimensional data can be difficult for contemporary systems. It is computationally costly, hard to interpret, and especially in the Machine Learning domain, it could lead the curse of dimensionality. The curse of dimensionality is a phenomenon that potentially arises when the data has too many variables and samples. Curse of dimensionality leads to poor predictive performance and high computational time required to train the predictive model [2].

There are some ways to avoid curse of dimensionality, one of them being performing relevant feature selection and class balance. By selecting only useful features and removing irrelevant or redundant attributes, the data processing could be more easier. Medical or clinical decision support system has often been defined as the complex process of

gathering, evaluating, analysing and interpreting medical data in order to formulate one or series of decision, judgement, or intervention [3]. Feature selection and extraction is not a new field within the Healthcare or Computer Science domain. the past, ever since the emergence of Knowledge Discovery from Databases (KDD) [4]. Data mining techniques have been used to find useful patterns in the data, to support medical decision making such as diagnosis process, choosing treatment option, and prognosis prediction. When applied properly, it could help the healthcare provider to improve patient's care [5]. However, medical data are typically complex and diverse, and discovering patterns in such kind of data using traditional method is difficult.

Coupled with the advancement in medical technology and digitalization of medical information, large amounts of patient-related medical data are being generated at a faster rate nowadays. Most dataset generated from newer technologies such as Polyethylene Terephthalate (PET) scan and Magnetic Resonance Imaging (MRI) are highly-dimensional [6]. There are also combinations of clinical and other types of data such as signal, ultrasound, and time-series data which increases data complexity and variability.

Feature selection is an important part of the data mining process, often performed as pre-processing step before applying data mining tasks, which aims to reduce dimensionality of data into a subset of relevant or optimal features. Even though there are several well-established ways to do it, there is no single 'best method' as each has their own advantages and disadvantages. The success in performing feature selection depends on what kind of metrics we want to use, how to carefully choose the proper method to achieve the goal, and understanding of the domain knowledge. Despite vast feature selection methods available, selecting the right feature subset can still be a challenging task. Generally, feature selection for medical predictive model follows either automated feature selection method, or relying on expert opinion [7].

Automated feature selection methods, from technical perspective, typically determine relevant features based on a set of mathematical rules [8] such as correlation between feature and target class, correlation between multiple features, or how each feature would accuracy, sensitivity, and specificity of certain classifier. Though promising, purely automated feature selection still poses some problems like computationally more expensive and increase in cost as the number of features grows. Other methods are faster but might have stability issues, such as when the same feature selection method is applied to the same data but produces different feature subsets with exactly the same performance. Data imbalance, one of the most common problem found in real medical data, could skew the result if not properly handled [9]. These problems can overcome by using Neural Network, the features are given as inputs and the risk is predicted as output as shown in Fig. 1.

On the other hand, medical experts do not rely on datasets and mathematical rules to select features. Through their existing domain knowledge, they can intuitively constrain knowledge discovery (thus avoid over fitting problems), describe relationships between attributes, categories of attributes, and correlation among them.

However, as humans have limited capability in processing huge amounts [10] of information, experts might not be able to make judgements as much and as fast as a computer does. Feature selection in medical data mining is still under active research, and new methods or combinations of existing methods are constantly developed. In order to
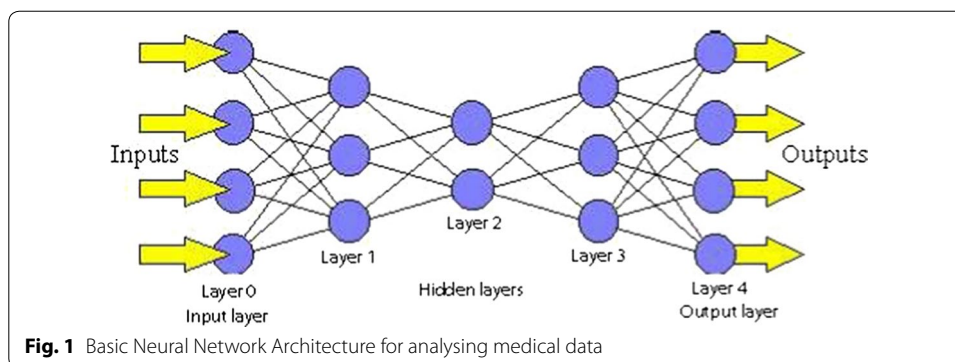
Fig. 1 Basic Neural Network Architecture for analysing medical data

solve such issues machine learning or deep learning is a best solution due to self-learned and more accurate in processing this kind of huge data. In this paper here we proposes an automated feature selection mechanism based on deep learning mechanism using MLP approach.

Related work

Based on the Dengue Haemorrhagic Fever (DHF) in Thailand has been examined a programmed expectation framework for DHF using entropy method and Artificial Neural Network (ANN) [11]. Entropy is utilized to extricate the significant data that influences the forecast precision. Afterward, the managed neural system is applied to anticipate future DHF flare-up. Result got uncovered that, by applying entropy method, it would yield a superior outcome as the entropy strategy produces 85.92% accuracy while just 78.16% when entropy not applied.

Wavelet change for information pre-handling before simulating Support Vector Machines (SVM)-based Genetic Algorithm in examining and anticipating the Dengue occurrence was proposed. The assessment and wellness of whole individual in the population is completed before entering next events. From their investigation, they found that, for predicting, Support Vector Regression (SVR) performed better contrasted with a direct degeneration and moreover increasingly reliable, even with the presence of overfitting [12].

Attribute Frequent Mining based on Dengue Outbreak was recognized a few quantities of credits to be utilized in deciding flare-ups as opposed to utilize the case studies. The relevant features utilized are year, week, age, sex, races, address, nature of work, kind of Dengue, hatching period, pandemic sort, intermittent cases and dead code and compared with Cumulative Sum (CUSUM) technique. The examination is focussed on utilizing different specific features dependent on Apriority idea has demonstrated a decent implementation of recognition rate, false positive rate and computational time. This analysis has been demonstrated through a test can be utilized to all the more likely distinguish flare-ups. The general consequences of the investigation found that the subsequent design can defeat the Cumulative Sum (CUSUM) in identifying the drawbacks [13].

A preceptive model [14] for disease location operating Ensemble Rule Based Classifiers was proposed which uses Decision Tree, Rough Set Classifier, Naïve Bayes, and

Associative Classifier. The different classifiers provide better accuracy up to 70% with the guidelines of a higher competence than effective single classifier.

Castro et al. [15] studied the Negative Selection Algorithm (NSA) which depends on the Artificial Immune Systems for discovery of Dengue episodes. Among the crucial points were their capacities to demonstrate conditions among traits by limiting assumptions about the basic attribute prediction. The significant inhibitions that were examined incorporate being a 'discovery' which produces result using a hidden model that is hard for a human to translate; that they are hard to use in the field because of the high computational cost of preparing; and that they are persuaded to over fitting the training information. Unique in relation to both neural system and decision tree strategies, Dengue flare-up can likewise be applied to the next procedure is Rough Set Theory (RST) that methodologies of upper and lower estimate of the information referenced that contributes from the pair of vague and uncertain information or known as unpleasant set [16].

Manogaran et al. [17] uses Hidden Markov Model (HMM) for information order demonstrates a procedure that changes with time. It can be viewed as a doubly inserted stochastic process with a procedure that is not perceptible and must be seen through another stochastic procedure that creates the interval of observations.

This method is said to possess greater advantages in which it finds the significance and hidden data, generate minimal of rules, find minimal set of data and easy to interpret; as what the current trend are focusing on such as in medical analysis, finance, banking and other fields including predicting the Dengue outbreak [18].

Electronic Health Record (HER) information mining which intends to predict the future data as indicated by the chronicled records from medical repository [19]. Med-2Vec [20] is the primary strategy to become familiar with the interpretable embedding's of clinical codes; however it supervises conditions of clinical codes among stopovers. Hold [21] is the principal interpretable model to scientifically ascertain the commitment of every clinical code to the present forecast by utilizing an opposite time consideration instrument in a Recurrent Neural Network (RNN) for paired expectation task. Dipole [22] is the principal work to embrace Bidirectional Repetitive Neural Systems (BRNN) and distinctive consideration components to improve the expectation precision. GRAM [23] is the main work to apply diagram put together consideration instrument with respect to the given clinical philosophy to learn strong clinical code embedding's in any event, when lack of training information and a RNN is utilized to demonstrate quiet stopovers. KAME utilizes elevated level information to improve the prescient presentation, which is expanding upon GRAM.

Proposed work

Neural Network is a nature-inspired predictive algorithm which is rooted in statistical technique similar to Logistic Regression, designed to mimic the workings of human brain, and contains series of mathematical equations used to simulate biological process such as learning and memory. ANN structure consists of simple, highly interconnected neurons, which are analogous to neurons in the brain. It receives a number of input signals and produces single output signal, which can be transmitted into many branches, and ends at incoming connections of other neurons in the network.

MLP is a feed forward Neural Network which consists of at least an input layer, one or more hidden layer, and an output layer, where each layer has its own specific function. Each of the hidden and output layer contains a number of neuron with activation function. By carefully constructing Neural Network architecture, such as the choice of activation function, or number of hidden layer and neuron in each layers, it can be used to model complex non-linear relationship between input variables and the outcome. Designing a good NN architecture for a given problem is not trivial. There is no general way to determine what is the good number of neuron and layers for Neural Network, thus the optimal structure is typically evaluated through experiments or experience of similar problem.

While Logistic Regression is considered as a white-box model, Neural Network (NN) is known as black-box model, means it does not allow interpretation of model parameters. Unlike Logistic Regression in which variable coefficient can be explicitly known, it is difficult to know what is happening on each neuron at each training iteration. However, due to its edibility and often high discriminative power, NN as predictive model is popular in domains where classification performance is more important than model interpretability.

Here the proposed work takes Medical data as input in a randomized portion of seventy percent of the data for training and remaining part for testing of the model. The proposed model extracts the relevant features and predicts the risk factor as shown in Fig. 2.

Input: medical data set

Output: classified medical data for prediction

Step-1: patronized the input data into two parts 70% of data for training and 30% for testing.

Step-2: feature extraction from data by using MLP

For (i=1 to n)

{

      Take each record from the dataset

      Apply MLP for extract the features

}

Step-3: feature selection from the extracted data

For (i=1 to n)

{

      Take each feature from the extracted features set

      Select the relevant features from step-2

}

Step-4: Train the model MLP from step-2 and 3.

Step-5: test the model with rest of the data

Proposed algorithm: MLP based Data classification

MLP for feature extraction and classification

Input layer is not count into number of layer, therefore the regular ANN is 2-layer NN or one-hidden-layer MLP. Formally, a 2-layer NN is a function f: $\triangleleft$D !$\triangleleft$L, where D is the size of the input vector x and L is size of the output vector f(x). In matrix notation with bias vectors b(1), b(2), weight matrices W(1) and W(2) and activation functions G and g.

$$f(x) = G\left(b^{(2)} + \mathbf{W}^{(2)}\left(g\left(b^{(1)} + \mathbf{W}^{(1)}x\right)\right)\right) \tag{1}$$

The activation function has to meet some requirements due to the learning process. It has to be non-constant, bounded, and monotonically-increasing continuous function. Using this function is guaranteed that the derivation exists and can be used during network learning. There are some base activations function for learning NN—Sigmoid, Tanh, Rectified Linear Unit (ReLu) and Maxout (Fig. 2).

The Sigmoid function takes a real-valued number and transforms it into the range between 0 and 1—in general, large negative numbers become 0 and large positive numbers become 1. The disadvantages of sigmoid function are that it saturates during learning and is not zero-centred. The Tanh function transforms a real-valued number to the range between -1 and 1. Like the sigmoid neurone, its activations saturate, but unlike the sigmoid neurone, its output is zero-centred. That is the reason why the Tanh is preferred to the Sigmoid. The ReLu is linear and non-saturating and simply thresholder negative values at zero. The advantages of ReLu are great acceleration the convergence compared to the Sigmoid and Tanh and it is easy to compute. The disadvantage is that ReLu can be deactivated during learning due to large gradient update. The last mentioned function is Maxout which is quite different because doubled ReLu function (the ReLu is the special type of Maxout function, where one input is always zero). This function benefits of ReLu and is not deactivated during learning. However, the disadvantage is that it produce two times more parameters, which have to be optimised.

To train the net, all parameters have to be optimised. The parameters to learn is the set = {W(2), b(2),W(1), b(1)}. The gradients @`/@can be obtained using the backpropagation algorithm—the weights are updated according to the gradient. The gradient is propagated from output to input, computed using loss function `(W,B|j) for all weights W and biases B in the net. For NN classification is the loss function for example:

$$\ell(\mathbf{W}, \mathbf{B}|\mathbf{j}) = \sum_{\hat{y}\in 0} y^{(j)} log\left(\hat{y}^{(j)}\right) + log\left(1 - \hat{y}^{(j)}\right)\left(1 - y^{(j)}\right) \tag{2}$$

where $\hat{y}$ is outputs from output layer and y is true class. NNs could have only one hidden layer and be universal approximator but they can also consist of many layers with many units. Even though the representational power is equal one hidden layer NN and two hidden layer NN, the deeper NN can provide better prediction because the structure is more complex and can reflect the real data better, but it is just an empirical observation.

The most commons are L1 and L2 regularisation. L1 means that each weight w is changed to |w|, where is regularisation strength—it leads the weight vectors to become sparse during optimisation, a sparse subset of the most important inputs is used and units become nearly invariant to the noisy inputs.

Fig. 2 Proposed architecture for analysing medical data
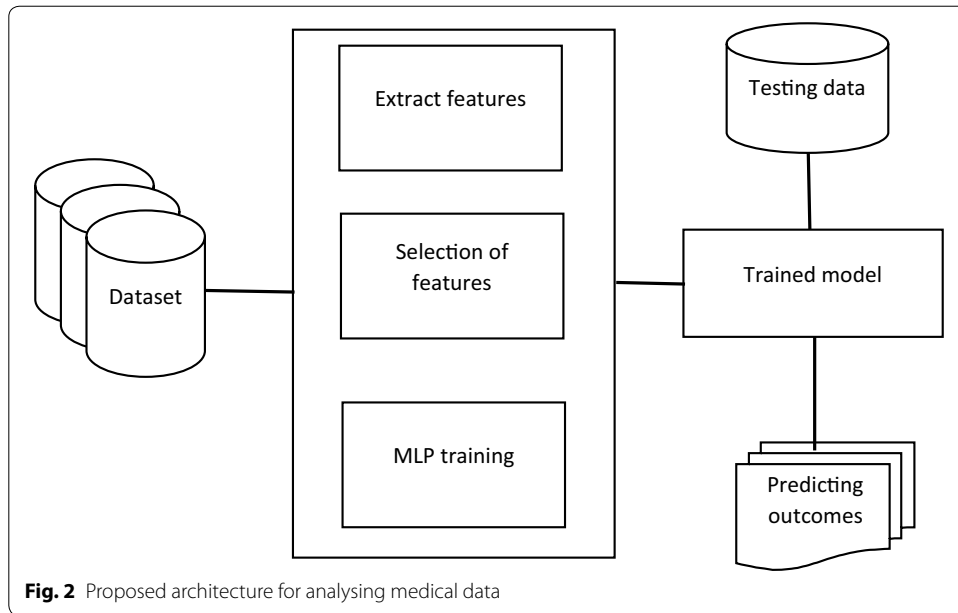
L2 means for every weight w in the network is changed to 1 2w2 where is regularisation strength again—it heavily penalise peak weights and prefers diffuse weight vectors, it allows the network to use all of its inputs a rather that some of its inputs a lot. These two regularisations could be combined. The hyper parameter is usually a small number, for example, $0 - 3$. Finally, the result function to optimise is:

$$\ell'(\mathbf{W}, \mathbf{B}|\mathbf{j}) = \ell(\mathbf{W}, \mathbf{B}|\mathbf{j}) + \lambda_1 L1(\mathbf{W}, \mathbf{B}|\mathbf{j}) + \lambda_2 L2(\mathbf{W}, \mathbf{B}|\mathbf{j}) \tag{3}$$

The Dropout is also very effective regularisation technique. While training, dropout is implemented by only keeping a neurone active with some probability p, or setting it to zero otherwise. Dropout can be interpreted as sampling a NN within the full-connected NN, and only updating the parameters of the sampled network based on the input data when the net is trained. During testing, there is no dropout applied—the prediction is evaluated across the exponentially-sized ensemble of all sub-networks. The hyper parameter p is a value between 0 and 1, for example 0.2.

$$f[m, n] * g[m, n] = \sum_{j=1}^{J} \sum_{i=1}^{I} f[i, j] g[m - i, n - j] \tag{4}$$

Experimental setup

All experiments have been conducted on a processor with Intel Core i5 having 16 GB of RAM and running the UBUNTU operating system. For applying various classification algorithms, here we used python 3.5 and tensor flow packages. We evaluate the performance of the models using different popular metrics from the literature, namely, accuracy, sensitivity (or recall), precision, F-score.

Datasets

The datasets we experimented in this paper are taken from University of California at Irvine (UCI) ML Repository. In specific we use 3 standard datasets, corresponding to three different diseases as termed below. Subsequently to do an autonomous testing of proposed model, intended for each disease, the samples for training and testing are separated.

The first dataset here we considered is a Wisconsin Breast Cancer (WBC) and it has total nine features and almost 500 samples are for training and 200 were testing samples. The second dataset we considered was SaHeart (SHt) data set and it contains total number of features as 9, and considered training samples are 300 and testing samples are almost 160 and the another dataset is Pima Indians Diabetes (PID) which contains 8 features in total and training samples are 576 and testing samples are 192.

The parameters considered in the dataset are represented in Table 1.

The first 30 features were computed from a input image of and all the specifications of the parameters are represented in Table 1.

Evaluation metrics

Evaluation metrics is a key factor in assessing performance of classifier. In two-class problem, confusion matrix is used to describe classifier predictive performance. Confusion matrix is a two-dimensional table that visualizes correctly or incorrectly predicted sample of each class by matching the actual and predicted value of all samples. It is especially useful in supervised learning, where each sample is labelled with a target class. In this example, actual class refers to the actual target class of an input data which is taken as the ground truth, while predicted class refers to classifier output of the same input data. In each dimension, label P denotes positive class and N denotes negative class. In reality, we can construct our own definition of positive and negative class, or go with common sense of what considered as positive and negative class. When prediction outcome of a classifier matches the actual class, we call such case True Positive if the actual and predicted class is positive, True Negative when actual and predicted class is negative. In case a classifier incorrectly predict negative class as positive or incorrectly predict positive class as negative, such case is termed False Positive or False Negative, respectively. For our CRT case study, we define our confusion matrix terms as follows:

Positive case: Patients has _LV EF $\geq$ 15%, considered as responsive to the treatment, labelled as class 1 in dataset.

Negative case: Patients has _LV EF < 15%, considered as unresponsive to CRT Treatment, labelled as class 0 in dataset.

Once confusion matrix is constructed, several other metrics can be derived from it, namely: Accuracy. Accuracy is the most common metrics used in data mining studies. It is simply defined as the proportion of correct decision made and total number of decisions made. However, accuracy does not always work well in all cases. In data with imbalance class, accuracy is no longer a proper measure because it could be biased to majority class and does not react the performance of minority class. It is possible that a classifier have a high accuracy by predicting all samples as negative case.

Table 1 Attribute information of the dataset

```
RangeIndex: 569 entries, 0 to 568
Data columns (total 31 columns):
diagnosis                  569 non-null object
radius_mean                569 non-null float64
texture_mean               569 non-null float64
perimeter_mean             569 non-null float64
area_mean                  569 non-null float64
smoothness_mean            569 non-null float64
compactness_mean           569 non-null float64
concavity_mean             569 non-null float64
concave points_mean        569 non-null float64
symmetry_mean              569 non-null float64
fractal_dimension_mean     569 non-null float64
radius_se                  569 non-null float64
texture_se                 569 non-null float64
perimeter_se               569 non-null float64
area_se                    569 non-null float64
smoothness_se              569 non-null float64
compactness_se             569 non-null float64
concavity_se               569 non-null float64
concave points_se          569 non-null float64
symmetry_se                569 non-null float64
fractal_dimension_se       569 non-null float64
radius_worst               569 non-null float64
texture_worst              569 non-null float64
perimeter_worst            569 non-null float64
area_worst                 569 non-null float64
smoothness_worst           569 non-null float64
compactness_worst          569 non-null float64
concavity_worst            569 non-null float64
concave points_worst       569 non-null float64
symmetry_worst             569 non-null float64
fractal_dimension_worst    569 non-null float64
dtypes: float64(30), object(1)
memory usage: 137.9+ KB
```

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

Precision: Precision (or also called Positive Predictive Value) is defined as proportion of positive case that are correctly identified in outcomes.

$$Precision = \frac{TP}{TP + FP}$$

Recall: Recall (also called Sensitivity or True Positive Rate) is defined as proportion of actual positive case which are correctly identified as positive case. High Recall value indicates low proportion of False Negatives within all positive samples. On the contrary, low Recall value is an indicator of higher False Negative proportion. Recall can also be interpreted as to which degree the positive samples are misclassified as negative case.

$$Recall = \frac{TP}{TP + FN}$$

Specificity: Specificity, also known as True Negative Rate is defined as proportion of actual negative case which is correctly identified as negative case. High Specificity value indicates low proportion of False Positives within all negative samples. On the contrary, low Specificity value is an indicator of higher False Positive proportion. Specificity can also be interpreted as to which degree the negative samples are misclassified as positive case.

$$Specificity = \frac{TN}{TN + FP}$$

F1 Score: F1 score is defined as weighted average of Precision and Recall.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Results and discussions

The subsequent section discuss about the performance of proposed mechanism with respect to different data sets which are taken from the UCI repository. The ANACONDA Platform is used for implementation and coding is done in python.

Here Fig. 3 shows that computation time for training data set by varying number of medical input records from different data sets for existing work and proposed work. Figure shows the varying of computation time here we use MLP for training. While increasing the number of input data records it gets more knowledge about but it need more processing time. Here we take running time of mechanism that is for training and testing on standalone machine for to understand how our mechanism works. With respect to existing and if we use it on cluster it will reduced based on the cluster capacity.

Here Fig. 4 shows that computation time for testing data set by varying number of medical input records from different data sets for existing work and proposed work. Figure shows the varying of computation time here we use MLP for testing. While increasing the number of input data records it gets more knowledge about but it need more processing time.

Here Fig. 5 shows the performance of existing proposed mechanism with respect to different performance metrics of precession, f-score, recall and accuracy. Here accuracy of prospered method and precession are increased with respect to data size. And recall and f-measure are gradually reduced. This shows that the proposed mechanism works well with respect to varying size of the data. The accuracy of the proposed method compared to traditional classification methods are depicted in Fig. 6. The existing methods considered are SVM, KNN and K-means methods and the accuracy of these methods are low when compared to proposed MLP based data classification.

Here Fig. 7 Accuracy comparison with state of art methods. In general classification of medical data is a complex task. And medical data needs to be accurate classification then other things. Here the fig shows the comparison of proposed model with respect to state of art models of LSTM and RNN. Previously RNN was used for classification of
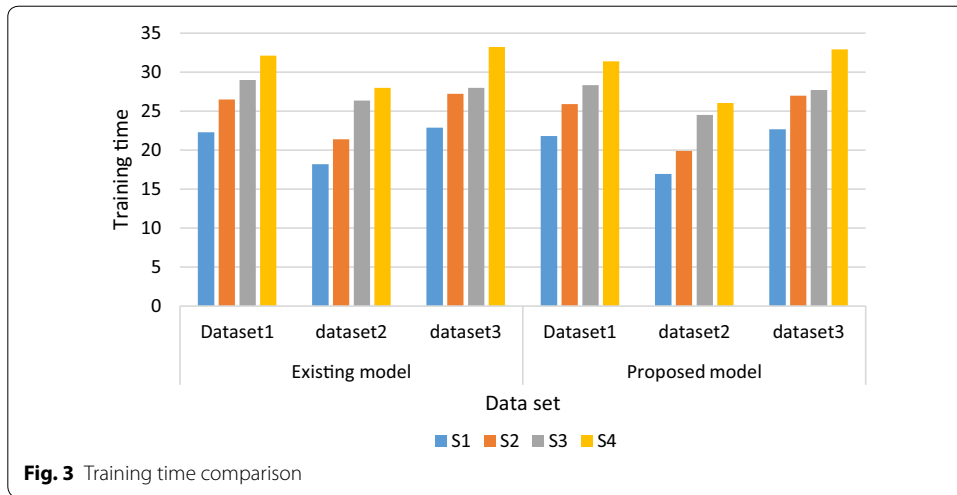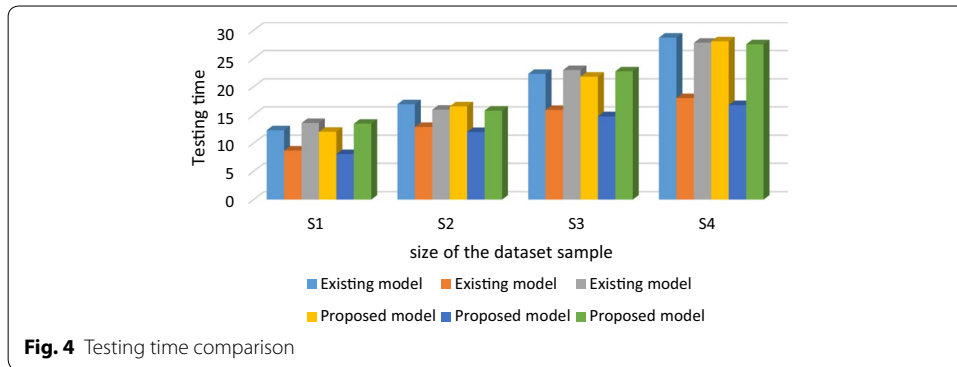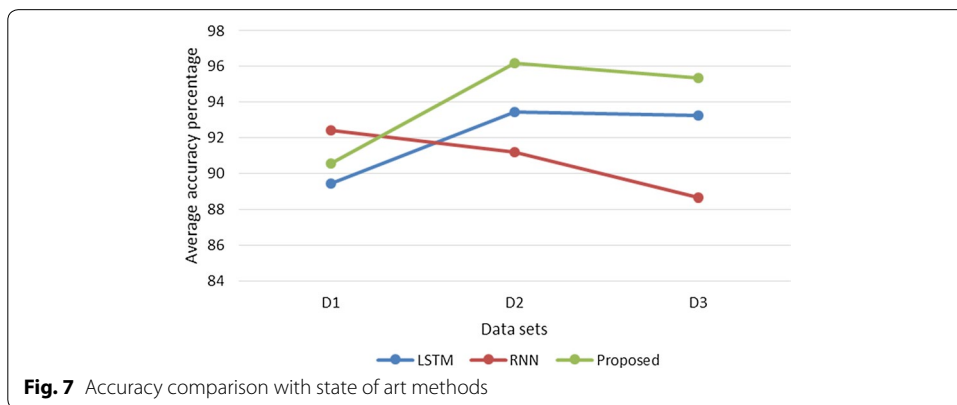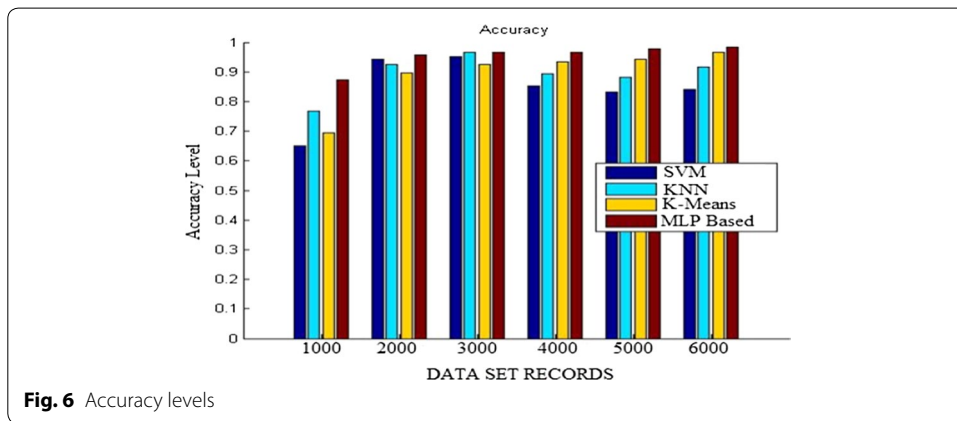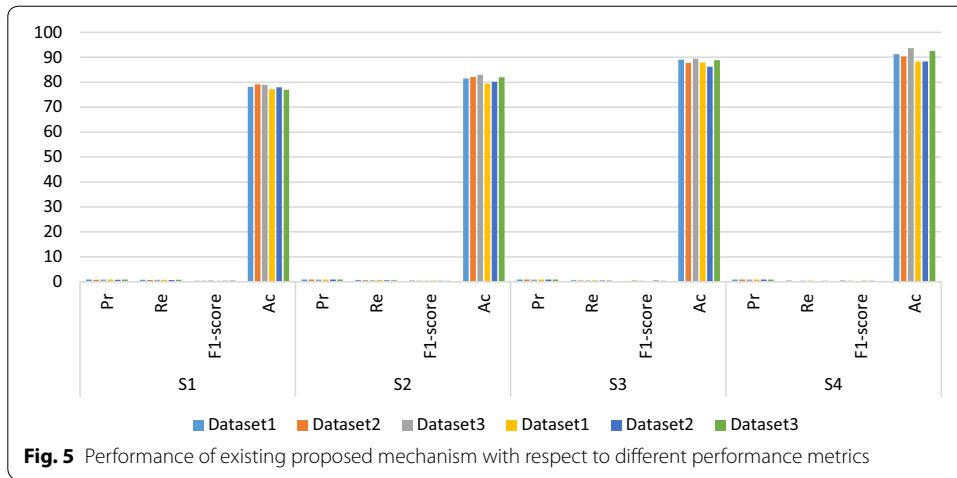
Fig. 3 Training time comparison

Fig. 4 Testing time comparison

medical data. But RNN has a drawback of it needs more training data for different kinds of issues. But medical data is a highly imbalanced so RNN fails most of the aspects in accurate classification. And LSTMs take more time for gets training, require huge memory and gets over fit easily. Medical data is huge making training of LSTM it occupies huge data. And more importantly LSTM needs more data to make accurate predictions. Previously we discussed medical data is imbalanced. So Proposed MLP based mechanism gives better classification accuracy than existing.

Conclusion

This paper mainly focuses on the deep learning based mechanism to predict the disease based on previous medical data. Here we used MLP for predicting the deceases and here we use feature extraction and classification of medical data. Lesser number of feature does not always produce higher performance. This is reflected in the result of MLP Neural Network classifier, in which distribution of selected feature subset number is low, yet also has lower overall performance. However, sometimes different number of feature might result in similar performance, as in case of state-of art mechanisms. Both cases are present in our experimental result. In future nested networks concept can be used to gain better nonlinear high-level features for representations of medical images, which

Fig. 5 Performance of existing proposed mechanism with respect to different performance metrics

Fig. 6 Accuracy levels

Fig. 7 Accuracy comparison with state of art methods

may achieve better performance than our model. Multi feature selection methods can be used in future to improve accuracy rate.

Abbreviations
MLP: Multi-layer perceptron; EHR: Electronic Health Record; KDD: Knowledge Discovery from Databases; PET: Polyethylene Terephthalate; MRI: Magnetic Resonance Imaging; ANN: Artificial Neural Network; SVM: Support Vector Machines;

CUSUM: Cumulative Sum; NSA: Negative Selection Algorithm; RST: Rough Set Theory; RNN: Recurrent Neural Network; BRNN: Bidirectional Repetitive Neural Systems; NN: Neural Network; ReLu: Rectified Linear Unit; WBC: Wisconsin Breast Cancer; UCI: University of California at Irvine.

References
1. Bikku T, Nandam SR, Akepogu AR. A contemporary feature selection and classification framework for imbalanced biomedical datasets. Egypt Inf J. 2018;19(3):191–198. https://doi.org/10.1016/j.eij.2018.03.003.
2. Carrasquilla J, Melko RG. Machine learning phases of matter. Nat Phys. 2017;13(5):431–434. https://doi.org/10.1038/nphys4035.
3. Fieschi M, Dufour JC, Staccini P, Gouvernet J, Bouhaddou O. Medical decision support systems: old dilemmas and new paradigms? Methods Inf Med. 2003;42(03):190–198. https://doi.org/10.1055/s-0038-1634394.
4. Liu, H., Motoda, H. Feature selection for knowledge discovery and data mining (Vol. 454). Springer Science & Business Media, 2012.
5. Helitzer DL, LaNoue M, Wilson B, de Hernandez BU, Warner T, Roter D. A randomized controlled trial of communication training with primary care providers to improve patient-centeredness and health risk communication. Patient Educ Couns. 2011;82(1):21–29. https://doi.org/10.1016/j.pec.2010.01.021.
6. Elhawary H, Zivanovic A, Davies B, Lamperth M. A review of magnetic resonance imaging compatible manipulators in surgery. Proceedings ofthe Institution of Mechanical Engineers, Part H: J Eng Med. 2006;220(3):413–424. https://doi.org/10.1243/09544119JEIM73.
7. Stoean R, Stoean C. Modeling medical decision making by support vector machines, explaining by rules of evolutionary algorithms with feature selection. Expert Sys Appli. 2013;40(7):2677–2686. https://doi.org/10.1016/j.eswa.2012.11.007.
8. Kotsiantis SB, Zaharakis I, Pintelas P. Supervised machine learning: A review of classification techniques. Emerg Arti Intell Appli Comput Eng. 2007;160(1):3–24.
9. Liu TY. Easyensemble and feature selection for imbalance data sets. In: 2009 international joint conference on bioinformatics, systems biology and intelligent computing, IEEE.2009. pp. 517–520. https://doi.org/10.1109/IJCBS.2009.22.
10. Deo RC. Machine learning in medicine. Circulation. 2015;132(20):1920–1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593.
11. Gubler DJ. Dengue/dengue haemorrhagic fever: history and current status. In: Novartis foundation symposium, Vol. 277, Chichester; New York; John Wiley; 1999. p. 3. https://doi.org/10.1002/0470058005.ch2.
12. Khelif R, Chebel-Morello B, Malinowski S, Laajili E, Fnaiech F, Zerhouni N. Direct remaining useful life estimation based on supportvector regression. IEEE Transactions on industrial electronics. 2016;64(3):2276–2285. https://doi.org/10.1109/TIE.2016.2623260.
13. Lai HW, Wang CC, Lai YC, Chen CJ, Lin SL, Chen ST, Kuo SJ. The learning curve of robotic nipple sparing mastectomy forbreast cancer: an analysis of consecutive 39 procedures with cumulative sum plot. Euro J Surg Oncol. 2019;45(2):125–133. https://doi.org/10.1016/j.ejso.2018.09.021.
14. Tran CT, Zhang M, Andreae P, Xue B, Bui LT. An ensemble of rule-based classifiers for incomplete data. In: 2017 21st Asia Pacific Symposium on Intelligent and Evolutionary Systems (IES), IEEE. 2017. pp. 7–12. https://doi.org/10.1109/IESYS.2017.8233553.
15. de Lima Costa JC, de Castro LN, de Paula Bianchini C. Sensitivity Analysis of the Negative Selection Algorithm Applied to Anomalies Identification in Builds. In: 2019 XLV Latin American Computing Conference (CLEI), IEEE, 2019. pp. 1–6. https://doi.org/10.1109/CLEI47609.2019.235087.
16. Bikku T, Nandam SR, Akepogu AR. An iterative Hadoop-based ensemble data classification model on distributed medical databases. In: Proceedings of the First International Conference on Computational Intelligence and Informatics. Springer, Singapore.2017. pp. 341–351. https://doi.org/10.1007/978-981-10-2471-9_3.

17. Manogaran G, Vijayakumar V, Varatharajan R, Kumar PM, Sundarasekar R, Hsu CH. Machine learning based big data processingframework for cancer diagnosis using hidden Markov model and GM clustering. Wireless Pers Commun. 2018;102(3):2099–2116. https://doi.org/10.1007/s11277-017-5044-z.
18. Rahmawati D, Sarno R, Fatichah C, Sunaryono D. Fraud detection on event log of bank financial credit business process using Hidden Markov Model algorithm. In: 2017 3rd International Conference on Science in Information Technology (ICSITech), IEEE.2017. pp. 35–40. https://doi.org/10.1109/ICSITech.2017.8257082.
19. Dhingra K, Panchal K. Efficient Mining of Cloud Based Electronic Health Records (EHR) for Clinical Decision Support System. Complexity, 2017; 2(1).
20. Yuan Y, Xun G, Suo Q, Jia K, Zhang A. Wave2vec: Learning deep representations for biosignals. In: 2017 IEEE International Conference on Data Mining (ICDM) IEEE. 2017. pp. 1159–1164. https://doi.org/10.1109/ICDM.2017.155.
21. Ma F. Deep Predictive Models for Mining Electronic Health Records (Doctoral dissertation, State University of New York at Buffalo), 2019.
22. Song L. Medical concept embedding with ontological representations. 2019. https://doi.org/10.24963/ijcai.2019/641.
23. Zhang X, Qian B, Li Y, Yin C, Wang X, Zheng Q. KnowRisk: an Interpretable Knowledge-Guided Model for Disease RiskPrediction. In: 2019 IEEE International Conference on Data Mining (ICDM), IEEE. 2019. pp. 1492–1497. https://doi.org/10.1109/ICDM.2019.00196.

Publisher's Note