IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 28, 2019, accepted November 21, 2019, date of publication November 29, 2019, date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2956859

A Multi-Task Neural Network Architecture for Renal Dysfunction Prediction in Heart Failure Patients With Electronic Health Records

BINHUA WANG[1,2,3], **YONGYI BAI**[1,4], **ZHENJIE YAO**[5,6], **JIANGONG LI**[5,6], **WEI DONG**[1,7], **YANHUI TU**[5,6], **WANGUO XUE**[3], **YAPING TIAN**[1,2], **YIFEI WANG**[3], **AND KUNLUN HE**[1,2,8]

[1]Beijing Key Laboratory of Precision Medicine for Chronic Heart Failure, Chinese PLA General Hospital, Beijing 100853, China
[2]Laboratory of Translational Medicine, Chinese PLA General Hospital, Beijing 100853, China
[3]Medical Big Data Center, Chinese PLA General Hospital, Beijing 100853, China
[4]Department of Cardiology, Second Medical Center, Chinese PLA General Hospital, Beijing 100853, China
[5]Purple Mountain Laboratory for Network Communication and Security, Nanjing 211111, China
[6]Rhinotech, Beijing 100124, China
[7]Department of Cardiology, Chinese PLA General Hospital, Beijing 100853, China
[8]Key Laboratory of Biomedical Engineering and Translational Medicine, Chinese PLA General Hospital, Beijing 100853, China

Corresponding authors: Zhenjie Yao (yaozhenjie@gmail.com), Wei Dong (301dongw@sina.com), and Kunlun He (kunlunhe@plagh.org)

The work of Y. Bai was supported by the Medical Big Data Project of PLA General Hospital under Grant 2018MBD-026. The work of W. Dong was supported by the National Key Research and Development Program of China under Grant 2018YFC0910700. The work of Y. Tian was supported by the National Key Research and Development Program of China under Grant 2017YFC1001700. The work of K. He was supported by the National Major Scientific and Technological Special Project for Significant New Drugs Development of China under Grant 2018ZX09J18109-004.

ABSTRACT Renal dysfunction, which is associated with bad clinical outcomes, is one of the most common complications of heart failure (HF). Timely prediction of renal dysfunction can help medical staffs intervene early to avoid catastrophic consequences. In this paper, we proposed a multi-task deep and wide neural network (MT-DWNN) for predicting fatal complications during hospitalization. The algorithm was tested on a dataset collected from Chinese PLA General Hospital, which contains 35,101 hospitalizations with HF diagnosis during the last 18 years, and 2,478 hospitalizations with a diagnosis of renal dysfunction. For the renal dysfunction task, the AUC of the proposed method is 0.9393, which is a significant improvement ($p < 0.01$) compared to that of conventional methods, while that of single task deep neural networks is 0.9370, that of random forest is 0.9360, and that of logistic regression is 0.9233. The experimental results show that the proposed MT-DWNN model achieves better prediction performance on renal dysfunction in HF patients than conventional models.

INDEX TERMS Heart failure, renal dysfunction, deep and wide neural network, multi-task.

I. INTRODUCTION

Heart failure (HF), also known as chronic heart failure (CHF), is a clinical syndrome caused by ventricular dysfunction. Heart failure is a serious manifestation or late stage of various heart diseases. Generally, heart failure would result in insufficient cardiac ejection. Heart failure is a disease with high mortality rate and high medical costs. According to the guidelines of European Society of Cardiology (ESC), there are about 26 million adults living with heart failure worldwide. The 5-year survival rate of HF is even worse than that of several kinds of cancer [1], [2]. HF was responsible for an estimated health expenditure of around $31 billion in the United States, which is equivalent to more than 10% of the total health expenditure for cardiovascular diseases [3].

As mentioned above, the immediate symptom of heart failure is insufficient blood ejection from the heart. Insufficient blood supply can cause diseases in other organs, such as kidney. Many previous researches have reached the same conclusion that renal dysfunction has a great adverse impact the prognosis of patients with HF [4], [5]. Bibbins-Domingo *et al.* showed that renal insufficiency was a major predictor of mortality among women with HF [6]. Forman *et al.* found that worsening renal function occurs frequently among hospitalized HF patients and is associated with significantly worse outcomes [7]. Hillege concluded that impaired renal function

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

is independently associated with heightened risk for death, cardiovascular death, and hospitalization for heart failure in patients with chronic HF with both preserved as well as reduced left ventricular ejection fraction [8]. Damman *et al.* found that worsening renal function predicts substantially higher rates of mortality and hospitalization in patients with HF [9]. Tobias *et al.* showed that worsening renal function was common in patients with acute decompensated heart failure and was linked to significantly worse outcomes. However, the clinical parameters failed to adequately predict its occurrence [10]. Cole *et al.* found that worsening renal function in heart failure is the result of a complex, multifactorial process [11].

These studies have found that worsening renal function is significantly associated with adverse outcomes in patients with heart failure. Take our dataset as an example: patients with renal dysfunction suffer higher risk of mortality. The hospital mortality rate of HF patients with renal dysfunction is 22.72%, compared to that of all the HF patient being 6.30%. Timely prediction of the renal dysfunction can help medical staff intervene early to avoid catastrophic consequences.

The remainder of this paper is organized as follows. Section II reviews related work on this topic.Section III introduces the dataset and the proposed model in detail. Section IV presents the experimental results on the dataset, comparison results of different neural networks and conventional machine learning models. Section V discusses the experimental results. Finally, Section VI gives conclusions and discusses future work.

II. RELATED WORKS

A. HEART FAILURE OUTCOMES PREDICTION

In the past few decades, many works have been dedicated to outcomes prediction of HF patients using machine learning (ML) algorithms. Guidi *et al.* presented a clinical decision support system (CDSS) for the analysis of heart failure (HF) patients, providing various outputs such as an HF severity evaluation, HF-type prediction, as well as a management interface that compares different patients' follow-ups. Compared to neural network (NN), support vector machine, fuzzy rules, and classification and regression tree, random forest obtained the best performance [12]. Masetic and Subasi compared many ML algorithms (decision tree, k-nearest neighbor, support vector machine, artificial neural networks, and random forest) on the task of detecting HF from ECG, and RF gives the best performance [13]. Rahimi *et al.* reviewed the literature for death and hospitalization risk prediction models in patients with HF [14]. Miao *et al.* proposed an improved random survival forest to predeict inhospital mortality [15]. Acharya *et al.* proposed a deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals [16]. Nirschl *et al.* developed a CNN classifier to detect HF from hematoxylin and eosin (H&E) stained whole-slide images [17]. Researchers also gave some other comprehensive surveys, which focus on applying machine learning to all aspects of the management of heart failure [2], [18]. However, most of these works focused on mortality or readmission prediction, with few focusing on complications of heart failure. Wosiak *et al.* proposed a multi-label classification technique for comorbidities identification, which is a general model for all kinds of diseases [19]. Xiang *et al.* proposed a multi-task framework that can jointly predict the risk of multiple related diseases, and the method was tested on patients with Congestive Heart Failure and Chronic Obstructive Pulmonary Disease [20].

B. DEEP LEARNING AND ITS APPLICATION IN HEALTHCARE

Deep learning (DL) is one of the most advanced machine learning models. In recent years, DL outperforms traditional machine learning methods with a significant margin in image recognition, speech recognition and natural language processing [21]. DL is also applied to solve many problems in healthcare [22]–[26]. Miotto *et al.* proposed an unsupervised representation of patient from the EHRs (Electronic Health Records), named Deep Patient, which is a three-layer stacked denoising autoencoders [27]. Cheng *et al.* represented the EHRs for every patient as a temporal matrix with time on one dimension and event on the other dimension. Then they built a four-layer convolutional neural network model to extract phenotypes and perform chronic diseases prediction [28]. Choi *et al.* used recurrent neural network models for early detection of heart failure onset [29]. Reference [22] and [30] for more works on applying DL to EHR data analysis.

In this paper, inspired by the wide & deep networks [31], we propose a novel deep and wide neural network, and apply it to predict renal dysfunction in HF patients with EHR data. The main contributions of the paper include:

- Applying deep learning to predict complication of HF patients.
- Proposing a novel deep and wide neural network architecture.
- Adding auxiliary task to improve the prediction performance.
- Validating the proposed model on a real world dataset.

III. MATERIALS AND METHODS

The immediate symptom of HF is insufficient blood ejection from the heart, which is prone to complications. Some serious complications, such as renal dysfunction are fatal. Timely prediction of the fatal complications can enable early intervention to avoid catastrophic consequences. Few works have been done on this topic. In this paper, we proposed a multi-task deep neural networks with novel architecture to improve prediction performance for renal dysfunction with electronic health records (EHRs). We also compared the proposed model with conventional machine learning methods.

A. DATASET AND PREPROCESSING

The dataset was extracted from Chinese PLA General Hospital (PLAGH), which contained Electronic Health

Records (EHR) data from January 2001 to August 2018. Based on the Hospital Information System (HIS) of PLAGH, which contained electronically available data from multiple institutional information systems contributing to the EHR, including demographic information, claims-based diagnosis and procedure information, laboratory values, and medication lists for all the patients hospitalized in PLGAH, a HF database was established. In the present study, patients were eligible for inclusion if they were: aged 18 years and above with a clinical diagnosis of HF according to the 2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure [1]. Patients were excluded if they decided to terminate in-hospital treatment, or were discharged/transferred to another short-term hospital or hospice. The final dataset included 35,101 records of 25,514 patients, including 15,694 males and 9,820 females. The average age of patients is 62.6($\pm$16.1) years old.

Among the 35,101 hospitalizations, 2,478 hospitalizations have a diagnosis of renal dysfunction, which are the positive samples for the renal dysfunction task. In addition, 2,211 patients died during hospitalization, resulting in the in-hospital mortality rate of 6.30%.

The study complies with the Declaration of Helsinki and has been approved by the Ethics Committee of the Chinese People's Liberation Army General Hospital.

A total of 89 features were adopted for prediction, including 4 basic indicators (gender, age, height, weight), 74 laboratory indicators (from blood gas, blood general form, carbohydrates and related substances, enzymes and other related substances, general urine test, lipids and their related substances, low molecular nitrogen compounds, pigments and related substances, protein and gelatin reaction and the electrolyte), 11 indicators from examination reports (Electrocardiogram (ECG) reports, Echocardiography reports ect., each indicates whether a patient has a certain type of symptoms, e.g., atrial fibrillation).

Each hospitalization is represented by 89 features extracted and preprocessed by the following steps:

- **First measurement extraction.** The aim of this step is to predict renal dysfunction at the very beginning of hospitalization, so only the first value of each item is used. If there are multiple duplicate values, the first one is selected. Some examinations in the outpatient clinic before hospitalization are taken into consideration if the gap between examination and hospitalization is less than 7 days. For example, single hospitalization may include 6 blood test results in total, but only the first result is taken. If there is a blood test three days prior to the hospitalization, we would take that one, instead of the first one during hospitalization.
- **Handling missing value.** For the missing features, a value of 0 was given to get the best performance after comparing several commonly used methods.
- **Normalization.** All the features are normalized by

$$x_{ij} = \frac{x_{ij} - \bar{x}_i}{std(x_i)} \tag{1}$$

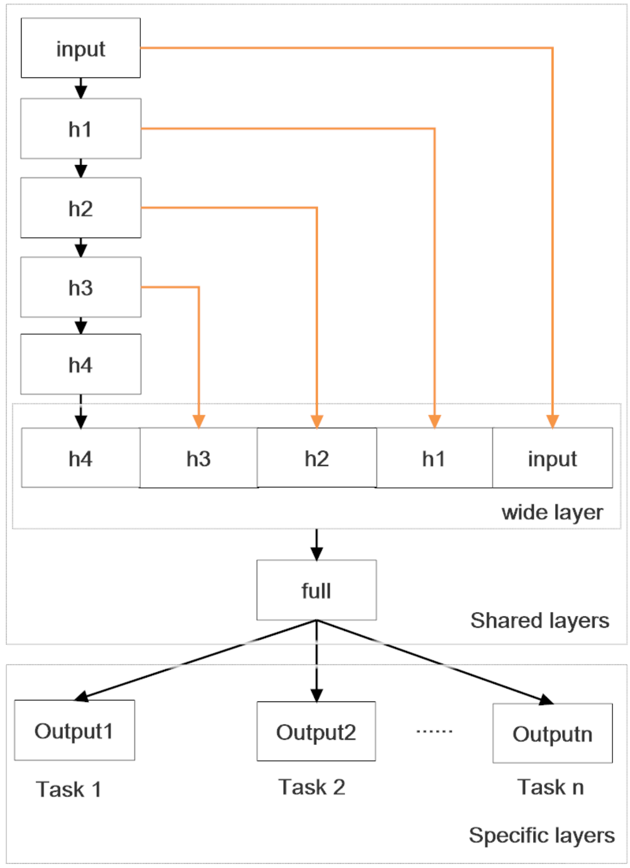


FIGURE 1. Overall architecture of MT-DWNN.

where x_{ij} is the value of the ith feature of the jth sample, x_i is the vector composed by ith feature of all the samples, $\bar{x}_i$ is the mean value of x_i, $std(x_i)$ is the standard deviation of x_i.

B. DWNN

In this paper, we present a novel architecture for deep neural networks named deep and wide neural networks (DWNN). Like other deep neural networks, DWNN is an end-to-end deep neural network model. Figure 1 presents the architecture of the whole Multi-task DWNN. The main improvement of DWNN focuses on the shared layer in the figure. The features extracted from EHR are fed into the DWNN as the input. As shown in the figure, the architecture is different from the conventional DNN. Similar to the conventional DNN, there are 4 hidden layers (labeled as hl,h2,h3,h4) in the DWNN. The main difference is that all the neurons in the input layer and all the hidden layers are concatenated in the wide layer. The wide layer is the input of the subsequent fully connected layers. The fully connected layers and output layer is similar to the conventional DNN.

As shown in Figure 1, compared to conventional DNN, we add a wide layer containing all the information in the input layer and all the hidden layers. Through the wide layer, direct connections (shortcuts) between all the hidden layers and fully connected layers are built (marked in orange in Figure 1).

Such direct connections shorten the distance between the output layer and all the intermediate layers. Take h2 for example, without the shortcuts, the only pathway from h2 to output is h2-h3-h4-full-output, which takes 4 steps. Shortcuts lead to 2 shortened pathways from h2 to output: one is h2-full-output; the other is h2-h3-full-output, which take 2 and 3 steps, respectively. The shortcuts tend to alleviate the vanishing-gradient problem, strengthen information propagation in both feed forward and back propagation procedure of the whole neural network.

In the training phase, the shortcuts could enhance the gradient back propagation. In the test phase, the wide layer contains information of all the forward layers with different depths (input, h1, h2, h3, h4). That is why the architecture adjustment is capable of improving the performance of conventional DNN.

C. MULTITASK NEURAL NETWORKS

With a focus on the prediction of renal dysfunction, this paper is aimed at the main task of predicting the occurrence of renal dysfunction during hospitalization. Although the prediction of only one complication is included, prediction of mortality and intubation have been adopted as auxiliary tasks (AT) to improve the performance, as they are the most commonly used outcomes of HF patient in hospitals.

The architecture of multi-task networks is shown in Figure 1. The input layer and all the hidden layers are shared layers, while the output layer is a specific layer for different tasks.

For multi-task neural networks, the whole network is trained simultaneously. As we know, the training procedure is supervised learning, so multi-task training leads to more supervision, which leads to more general representation of the data. Generally speaking, multi-task mechanism results in better performance and improves generalization ability.

D. MULTITASK DWNN AND ITS IMPLEMENTATION DETAILS

In this paper, to build a model for renal dysfunction prediction, the DWNN was modified for multi-task learning.

The input layer, hidden layers, the wide layer, and the fully connected layer of DWNN are all shared layers. Specifically, the input layer includes 89 neurons, each corresponding to 1 feature. The 4 hidden layers are with dense connections, and each layer includes 512 neurons. Each hidden layer is composed of dense connections, batch normalization, 'relu' activation, and dropouts. The dropout rate was set to 0.65. In the wide layer, the input layer and all the hidden layers are concatenated together and fed into the fully connected layers. The fully connected layer including 1024 neurons with the dropout rate of 0.65, but the activation is changed to 'sigmoid'.

The output layer of DWNN is changed to 3 independent softmax layers, corresponding to one main prediction task: renal dysfunction, and 2 auxiliary tasks: morality and intubation. For example, for the renal dysfunction task, the result should be positive if the patient was diagnosed with renal dysfunction during hospitalization. Conversely, the result should be negative if the patient was not diagnosed with renal dysfunction during the hospitalization. The codes of building the neural networks are available at https://github.com/yzjba/dwnn.

The network is trained using Adam, an efficient stochastic gradient-based optimizer [32]. The batch size is set to 1024.

IV. RESULTS

A. MEASUREMENTS

To measure the performance of the proposed method for complications prediction, sensitivity (Sen), specificity (Spec), Detection Rate (DR), False Alarm Rate (FAR) and accuracy (Acc) are calculated, which are defined as:

$$Sen = TP/(TP + FN), \tag{2}$$

$$Spec = TN/(TN + FP), \tag{3}$$

$$DR = Sen, \tag{4}$$

$$FAR = 1 - Spec, \tag{5}$$

$$Acc = (TP + TN)/(TP + FN + TN + FP), \tag{6}$$

in which TP (True Positive) is the number of positive samples that are recognized as positive; FN (False Negative) is the number of positive samples that are recognized as negative; TN (True Negative) is the number of negative samples that are recognized as negative; FP (False Positive) is the number of negative samples that are recognized as positive. In general, a good detection method will minimize the FAR while maximize all the other 4 performance metrics.

B. REFERENCE MODELS

For comparison, we also tested Logistic Regression (LR), Random Forest (RF), Multi-task DNN (MT-DNN), and Single task DWNN (ST-DWNN) on the same dataset. LR is the most commonly used statistical analysis model in medical community. RF is one of the most commonly used machine learning models. MT-DNN is adopted as reference model to show the effectiveness of the proposed DWNN architecture. ST-DWNN is adopted as reference model to show the effectiveness of multi-task learning.

In the MT-DNN, the wide layer and the direct connections are removed. The input layer, hidden layers, fully connected layer, and output layer are the same as MT-DWNN. The hidden layer h4 is directly connected to the first fully connected layer. The output tasks are renal dysfunction and auxiliary task prediction, respectively.

In the ST-DWNN, the input layer, hidden layers, wide layer, and fully connected layer are the same as MT-DWNN. No auxiliary task was added. The only task is renal dysfunction prediction.

C. PARAMETER SETTING

The dataset was segmented into training set and testing set. Data from 2001 to 2013 were used as training set (21,104 admissions) and data from 2014 to 2018 were

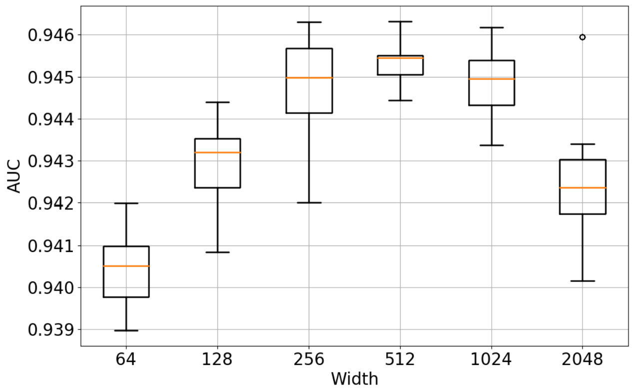


FIGURE 2. Comparison of different widths.

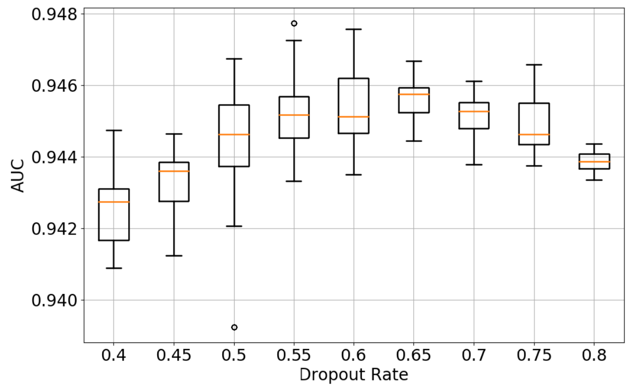


FIGURE 3. Comparison of different dropout rate.

used as testing set (13,997 admissions). All the parameters were determined by 3-fold cross-validation on training dataset. And the performance was tested on testing dataset.

All the experiments were run on a desktop with 4-core Intel(R) Core(TM) processor of 3.20 GHz, 8GB RAM. Training on 21,104 samples took 148.95s. Testing on 13,997 samples took 2.14s.

There are two important parameters in MT-DWNN. The first one is width (number of neurons per hidden layer) of the hidden layers and the second one is dropout rate.Differdent values have been tried for the two parameters.

A quantitative measurement of detection model is Area Under Curve (AUC), which is defined as ratio of area under the Receiver Operating Characteristic (ROC) curve. Generally, the larger the AUC is, the better performance the detector gets.

Figure 2 lists the AUCs of MT-DWNN with different width, where dropout rate is fixed to 0.65. As shown in the table, DenseDNN with 512 neurons per hidden layer perform the best. The median AUC is 0.9455. Figure 3 lists AUCs of MT-DWNN with different dropout rates, with width being fixed to 512. It can be found that MT-DWNN with a dropout rate of 0.65 gives the highest AUC of 0.9457. In the following tests, 512 is selected as the width, and 0.65 is selected as the dropout rate.

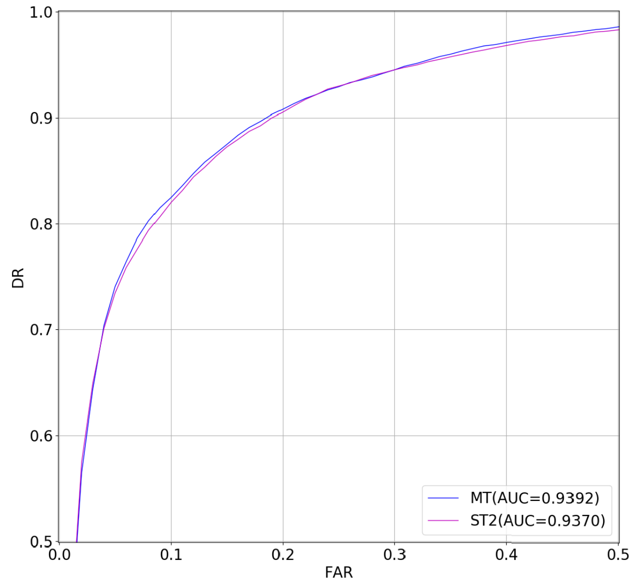


FIGURE 4. Renal dysfunction prediction performance of MT-DWNN and ST-DWNN.

D. MT-DWNN VS ST-DWNN

To show the effectiveness of multi-task learning, we compare MT-DWNN with single task networks (ST-DWNN).

ROC curves of MT-DWNN, ST-DWNN for renal dysfunction prediction are given in Figure 4. To facilitate the comparison, only the top-left part (the most important part) of the ROCs was shown. For each curve, with the increasing false alarm rate, the detect rate of the detectors rises up. It is found that, for most cases, the ROC curve of MT-DWNN is higher than that of the ST-DWNN, with a significant margin. As shown in Figure 4, the AUC of MT-DWNN is 0.9393 while the AUC of ST-DWNN is 0.9370. Repeated statistics of AUC shows that there is a significant difference between the groups of MT-DWNN and ST-DWNN ($p<0.001$, shown in Table 1).

According to the ROC curves and their AUCs, we concluded that, for the renal dysfunction prediction, the multi-task learning improves the performance significantly. Multi-task learning mechanism is very helpful for the complication prediction task.

E. MT-DWNN VS MT-DNN

To show the effectiveness of the DWNN architecture, we compared MT-DWNN with MT-DNN. As mentioned above, MT-DNN is a commonly used multi-task deep neural network.

ROC curves of MT-DWNN, and MT-DNN for renal dysfunction prediction are given in Figure 5. It is found that, for most cases, the ROC curve of MT-DWNN is higher than that of MT-DNN with a significant margin. The AUC of MT-DWNN for renal dysfunction prediction is 0.9393, while the AUC of MT-DNN is 0.9343. Repeated statistics of AUC shows that there is a significant difference between the groups of MT-DWNN and MT-DNN ($p<0.001$, shown in Table 1).

TABLE 1. Prediction performance of different models for renal dysfunction.

Model	Sen with different Specs					AUC	p-value vs MT-DWNN
	0.95	0.90	0.85	0.80	0.75		
LR	0.7012	0.8119	0.8559	0.8875	0.9060	0.9233(±0.0001)	<0.01
RF	0.7109	**0.8256**	0.8739	0.9018	0.9241	0.9360(±0.0006)	<0.01
ST-DWNN	0.7343	0.8199	0.8724	0.9050	**0.9295**	0.9370(±0.0013)	<0.01
MT-DNN	0.7192	0.8245	0.8729	0.9050	0.9269	0.9343(±0.0007)	<0.01
MT-DWNN	**0.7405**	0.8246	**0.8748**	**0.9077**	0.9289	**0.9392**(±0.0006)	1.0

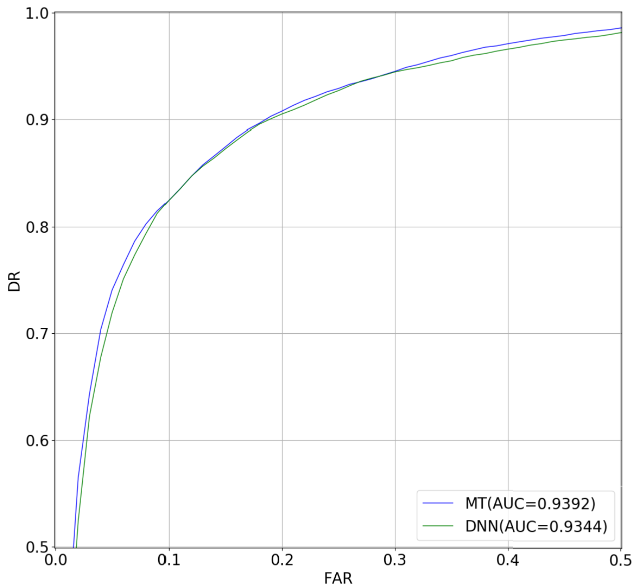


FIGURE 5. Renal dysfunction prediction performance of MT-DWNN and MT-DNN.

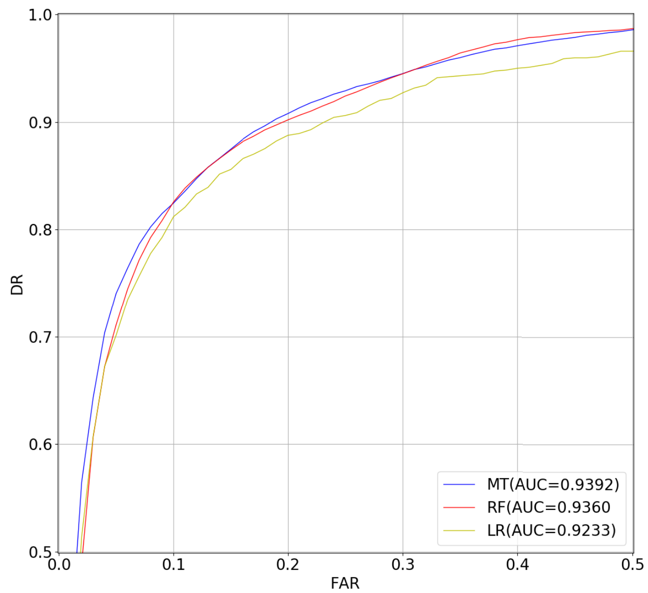


FIGURE 6. Renal dysfunction prediction performance of MT-DWNN and conventional models.

According to the ROC curves and their AUCs, we conclude that, for renal prediction, the DWNN architecture improves the performance. Adding direct connections improves the complication prediction performance.

F. MT-DWNN VS CONVENTIONAL METHODS

To further show the effectiveness of the MT-DWNN architecture, we compared MT-DWNN with two conventional machine learning methods, LR and RF. Both are the most commonly used machine learning models in the medical community.

ROC curves of MT-DWNN, LR, and RF for renal dysfunction prediction are given in Figure 6. We find that, the ROC curve of MT-DWNN is always higher than that of LR and RF. The AUC of MT-DWNN for renal dysfunction prediction is 0.9393, the AUC of RF is 0.9360, while the AUC of LR is 0.9233. Repeated statistics of AUC shows that there was a significant difference between the groups of MT-DWNN and LR, and that of MT-DWNN and RF (p<0.001, shown in Table 1).

According to the ROC curves and their AUCs, we concluded that, for renal dysfunction prediction, the MT-DWNN architecture improves the performance significantly than conventional machine learning methods. The proposed model improves the complication prediction performance.

The Spec, Sen, and AUCs of all the models for renal dysfunction prediction are listed in Table 1. 5 specificity values are selected as reference, which are 0.95, 0.90,0.85, 0.80, and 0.75. We can figure out that the proposed MT-DWNN achieves the largest sensitivity on 3 specificities and ST-DWNN and RF get 1 largest specificity. The AUC of MT-DWNN is 0.9393, which is larger than that of all the other models. The comparison shows that the proposed MT-DWNN is capable of predicting renal dysfunction in heart failure patients more effectively.

G. AUXILIARY TASK ANALYSIS

To check the contributions of the auxiliary tasks, we train the DWNN with different auxiliary tasks. The AUCs of different models are listed in Table 2. As shown in the table, the following conclusions can be reached:

Models with auxiliary tasks perform better than model without auxiliary tasks. The AUCs of DWNN with any kind and number outperform that of DWNN without auxiliary task significantly ($p < 0.01$ or $p < 0.01$).

Compared to the results of model with one auxiliary task, we find that the model with mortality as auxiliary task performs similar to the model with intubation as auxiliary task. Through checking the relationship between the main task and both ATs, it is found 22.72% patients with renal dysfunction died in hospital, and 17.55% received intubation treatment.

TABLE 2. AUCs of DWNN with different auxiliary tasks.

Model	No AT	AT=mortality	AT=intubation	Both ATs
DWNN	0.9370(±0.0013)	0.9389(±0.0010)	0.9386(±0.0011)	0.9392(±0.0006)
p-value (vs No ATs)	1.0	<0.01	<0.05	<0.01
p-value (vs Both ATs)	<0.01	>0.05	>0.05	1.0

The correlation between renal dysfunction and mortality is similar to that between renal dysfunction and intubation.

Model with both auxiliary tasks obtians the best performance, which is significantly better than DWNN without AT ($p < 0.05$), but not significant compared to DWNN with one AT ($p > 0.05$). Generally, model with more auxiliary tasks means more information is fed into the training procedure, hance resulting in better performance.

V. DISCUSSIONS

As shown in Figure 4, MT-DWNN outperforms ST-DWNN, the single-task baseline, on the prediction task. From Table 1, we can also figure out that, multi-task models, regardless of the type (mortality or intubation) and number (one or two) of auxiliary tasks, outperform ST-DWNN. Furthermore, auxiliary task selection is important for multi-task learning. The auxiliary task more relevant to the main task tends to get improved performance than the auxiliary task less relevant to the main task. Multi-task neural networks are a natural fit for complication risk prediction [20], [33]. Auxiliary tasks would be helpful even though the main aim is single task prediction. Herein, the training procedure of MT-DWNN takes all the information from all the 3 tasks (renal dysfunction, mortality, and intubation) into consideration, which leads to improved performance and better generalization ability.

Figure 5 and Table 1 showed that MT-DWNN achieved improved prediction performance compared to MT-DNN, thus proving the benefits of the DWNN architecture. The proposed DWNN jointly learns from information across all the layers. The shortcuts across the hidden and wide layers strengthen information propagation in both feed forward and back propagation procedures of the whole neural network, making the training and predicting procedure more effective.

In summary, it is found that the multi-task learning mechanism and the DWNN network architecture improve the performance for renal dysfunction prediction. But the improvement is limited. The main disadvantage is that, the architecture cannot capture the dynamically changing trend of the items in EHR. Further work would include readjusting the architecture of DWNN and changing the data representation for hospitalizations to capture the changing trend of items in EHR.

VI. CONCLUSION

In this paper, we proposed a multi-task deep and wide neural network to predict renal dysfunction in HF patients. It provides a basis for clinicians to timely individualize the risk of the renal function worsening in patients with HF so as to intervene early to avoid catastrophic consequences. The main task is to predict the occurrence of renal dysfunction during hospitalization. The model was tested on a dataset containing 25,514 heart failure patients during January 2001 to August 2018. The experimental results show that the proposed MT-DWNN model achieves better prediction performance than conventional models. By analyzing the results, it is found that the multi-task learning mechanism and the DWNN network architecture improve the performance for renal dysfunction prediction. Moreover, it is also found that the auxiliary task tends to obtain improved performance if it is more relevant to the main task. Future work will include incorporating more information in EHR into our framework and improving the architecture of the DNN, aiming to further improve the prediction performance.

ACKNOWLEDGMENT

(Binhua Wang and Yongyi Bai contributed equally to this work.)

REFERENCES

[1] P. Ponikowski *et al.*, "2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure," *Eur. J. Heart Failure*, vol. 18, no. 8, pp. 891–975, 2016.

[2] E. E. Tripoliti, T. G. Papadopoulos, G. S. Karanasiou, K. K. Naka, and D. I. Fotiadis, "Heart failure: Diagnosis, severity estimation and prediction of adverse events through machine learning techniques," *Comput. Struct. Biotechnol. J.*, vol. 15, pp. 26–47, Nov. 2017.

[3] G. Savarese and L. H. Lund, "Global public health burden of heart failure," *Cardiac Failure Rev.*, vol. 3, no. 1, pp. 7–11, 2017.

[4] S. S. Gottlieb, W. Abraham, J. Butler, D. E. Forman, E. Loh, B. M. Massie, C. M. O'Connor, M. W. Rich, L. W. Stevenson, and J. Young, "The prognostic importance of different definitions of worsening renal function in congestive heart failure," *J. Cardiac Failure*, vol. 8, no. 3, pp. 136–141, 2002.

[5] G. L. Smith, J. H. Lichtman, M. B. Bracken, M. G. Shlipak, C. O. Phillips, P. DiCapua, and H. M. Krumholz, "Renal impairment and outcomes in heart failure: Systematic review and meta-analysis," *J. Amer. College Cardiol.*, vol. 47, no. 10, pp. 1987–1996, 2006.

[6] K. Bibbins-Domingo, F. Lin, E. Vittinghoff, E. Barrett-Connor, D. Grady, and M. G. Shlipak, "Renal insufficiency as an independent predictor of mortality among women with heart failure," *J. Amer. College Cardiol.*, vol. 44, no. 8, pp. 1593–1600, 2004.

[7] D. E. Forman, J. Butler, Y. Wang, W. T. Abraham, C. M. O'Connor, S. S. Gottlieb, E. Loh, B. M. Massie, M. W. Rich, and L. W. Stevenson, "Incidence, predictors at admission, and impact of worsening renal function among patients hospitalized with heart failure," *J. Amer. College Cardiol.*, vol. 43, no. 1, pp. 61–67, 2004.

[8] H. L. Hillege, D. Nitsch, M. A. Pfeffer, K. Swedberg, J. J. Mcmurray, S. Yusuf, C. B. Granger, E. L. Michelson, J. Ostergren, and J. H. Cornel, "Renal function as a predictor of outcome in a broad spectrum of patients with heart failure," *Circulation*, vol. 114, no. 6, pp. 671–678, 2006.

[9] K. Damman, G. Navis, A. A. Voors, F. W. Asselbergs, T. D. Smilde, J. G. Cleland, D. J. van Veldhuisen, and H. L. Hillege, "Worsening renal function and prognosis in heart failure: Systematic review and meta-analysis," *J. Cardiac Failure*, vol. 13, no. 8, pp. 599–608, 2007.

[10] B. Tobias, S. Thenral, N. Markus, K. Theresia, H. Corinna, R. Tobias, P. Mihael, N. Albina, T. Christopher, and A. Nisha, "Effect and clinical prediction of worsening renal function in acute decompensated heart failure," *Amer. J. Cardiol.*, vol. 107, no. 5, pp. 730–735, 2011.

[11] R. T. Cole, A. Masoumi, F. Triposkiadis, G. Giamouzis, V. Georgiopoulou, A. Kalogeropoulos, and J. Butler, "Renal dysfunction in heart failure," *Med. Clinics North Amer.*, vol. 96, no. 5, pp. 955–974, 2012.

[12] G. Guidi, M. C. Pettenati, P. Melillo, and E. Iadanza, "A machine learning system to improve heart failure patient assistance," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 6, pp. 1750–1756, Nov. 2014.

[13] Z. Mašetic and A. Subasi, "Congestive heart failure detection using random forest classifier," *Comput. Methods Programs Biomed.*, vol. 130, pp. 54–64, Jul. 2016.

[14] K. Rahimi, D. Bennett, N. Conrad, T. M. Williams, J. Basu, J. Dwight, M. Woodward, A. Patel, J. Mcmurray, and S. Macmahon, "Risk prediction in patients with heart failure: A systematic review and analysis," *JACC Heart Failure*, vol. 2, no. 5, pp. 440–446, 2014.

[15] F. Miao, Y.-P. Cai, Y.-X. Zhang, X.-M. Fan, and Y. Li, "Predictive modeling of hospital mortality for patients with heart failure by using an improved random survival forest," *IEEE Access*, vol. 6, pp. 7244–7253, 2018.

[16] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, and R. S. Tan, "Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals," *Appl. Intell.*, vol. 49, no. 1, pp. 16–27, 2019.

[17] J. J. Nirschl, A. Janowczyk, E. G. Peyster, R. Frank, K. B. Margulies, M. D. Feldman, and A. Madabhushi, "A deep-learning classifier identifies patients with clinical heart failure using whole-slide images of H&E tissue," *PLoS One*, vol. 13, no. 4, pp. 1–16, 2018.

[18] S. E. Awan, F. Sohel, F. M. Sanfilippo, M. Bennamoun, and G. Dwivedi, "Machine learning in heart failure: Ready for prime time," *Current Opinion Cardiol.*, vol. 33, no. 2, p. 1, 2018.

[19] A. Wosiak, K. Glinka, and D. Zakrzewska, "Multi-label classification methods for improving comorbidities identification," *Comput. Biol. Med.*, vol. 100, pp. 279–288, Sep. 2018.

[20] W. Xiang, W. Fei, and J. Hu, "A multi-task learning framework for joint disease risk prediction and comorbidity discovery," in *Proc. 22nd Int. Conf. Pattern Recognit.*, 2014, pp. 220–225.

[21] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[22] Y. Zhen-Jie, B. Jie, and C. Yi-Xin, "Applying deep learning to individual and community health monitoring data: A survey," *Int. J. Autom. Comput.*, vol. 15, no. 6, pp. 643–655, 2018.

[23] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018.

[24] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[25] Z. Yao, Z. Zhu, and Y. Chen, "Atrial fibrillation detection by multi-scale convolutional neural networks," in *Proc. Int. Conf. Inf. Fusion*, 2017, pp. 1–6.

[26] Z. Yao and Y. Chen, "Arrhythmia classification from single lead ecg by multi-scale convolutional neural networks," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Jul. 2018, pp. 344–347.

[27] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep patient: An unsupervised representation to predict the future of patients from the electronic health records," *Sci. Rep.*, vol. 6, May 2016, Art. no. 26094.

[28] Y. Cheng, F. Wang, P. Zhang, and J. Hu, "Risk prediction with electronic health records: A deep learning approach," in *Proc. SIAM Int. Conf. Data Mining*, 2016, pp. 432–440.

[29] E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, "Using recurrent neural network models for early detection of heart failure onset," *J. Amer. Med. Inform. Assoc.*, vol. 24, no. 2, p. 361, 2017.

[30] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 5, pp. 1589–1604, Sep. 2018.

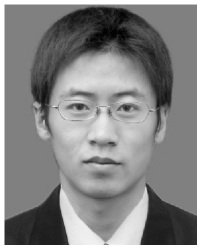
[31] H. T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai, and M. Ispir, "Wide & deep learning for recommender systems," in *Proc. 1st Workshop Deep Learn. Recommender Syst.*, 2016, pp. 7–10.

[32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[33] S. Ruder, "An overview of multi-task learning in deep neural networks," 2017, *arXiv:1706.05098*. [Online]. Available: https://arxiv.org/abs/1706.05098

BINHUA WANG received the Ph.D. degree in biomedical engineering from Shandong University. His research interests include smart instrument development, medical data mining analysis, and artificial intelligence.

YONGYI BAI received the Ph.D. degree. He is currently a Deputy Chief Physician. His research interests include diagnosis and management of cardiovascular diseases, such as hypertension, coronary heart disease, and heart failure.



ZHENJIE YAO received the B.S. degree in instrument science from Zhejiang University, China, in 2007, and the Ph.D. degree in communication engineering from the University of Chinese Academy of Sciences, in 2012. In 2012, he was a Researcher with the China Mobile Research Institute, China. He is currently a Data Scientist with Rhinotech LLC, Beijing, China. He has published about 20 refereed journal articles and conference papers. He holds 20 patents. His research interests include machine learning, mHealth, and signal processing.

JIANGONG LI received the M.S. degree from Tianjin University, in 2006. He was the Leader of the e-health Direction of China Unicom Research Institute. His research interests include the Internet of Things, cloud computing, big data, and artificial intelligence in the medical and health industry.

WEI DONG received the Ph.D. degree. She is currently a Chief Physician. Her research interests include diagnosis and treatment of chronic heart failure, cardiomyopathy, hypertension, hyperlipidemia, diagnosis of cardiomyopathy by cardiac NMR, and precise adjustment of cardiovascular drugs.

YANHUI TU received the B.S. degree in computer science from Beihang University, Beijing, China, in 2002. His research interests include cloud computing, big data, data mining, machine learning, and artificial intelligence.

WANGUO XUE is currently a Senior Engineer. His research interests include information system in hospital, system integration, and medical data mining.

YAPING TIAN received the Ph.D. degree. He is currently a Professor. His research interests include early laboratory diagnosis of major diseases, such as tumors, cardiovascular, and cerebrovascular diseases.

YIFEI WANG received the B.S. and M.Sc. degrees in biomedical engineering from Xi'an Jiaotong University, Xi'an, China, in 2013 and 2018, respectively. Her research interests include medical data mining and biomedical signal processing.

KUNLUN HE received the Ph.D. degree. He is currently a Chief Physician and a Professor. His research interests include diagnosis and treatment of cardiovascular diseases, such as hypertension, coronary heart disease, ischemic cardiomyopathy and heart failure, and translational medicine.