Hybrid Label Noise Correction Algorithm For Medical Auxiliary Diagnosis

Jiwei, Xu
Yunnan University
National Pilot School of software
Kunming, China
420076887@qq.com

Yun Yang
Yunnan University
National Pilot School of software
Kunming, China
yangyun@ynu.edu.cn

Po Yang
Sheffield University
Department of Computer
London , England
po.yang@sheffield.ac.uk

***Abstract*—In the context of the continuous development of Internet of Things (IoT) technology and Machine learning (ML) technology, its application in the medical field is becoming more and more extensive. However, with a dramatic increase in medical data obtained from the IoT-based medical auxiliary diagnosis system, the impact of label noise problems is also increasing. When training a machine learning algorithm for a supervised-learning task in some clinical applications, uncertainty in the labels of some patients may adversely affect the performance of the algorithm. For example, due to ambiguous patient conditions or poor reliability of diagnostic criteria, even clinical experts may lack confidence in making medical diagnoses for some patients. As a result, some samples used in algorithm training may be mislabeled, which adversely affects the performance of the algorithm. In this paper, we study a classification problem of sample labels with random damage. We propose a new hybrid label noise correction model that generalizes many learning problems, including supervised, unsupervised and semi-supervised learning. This hybrid model can withstand the negative effects of random noise and various non-random label noise. Extensive experimental results using real-world datasets from UCI machine learning repository are provided, the empirical study shows that our approach successfully improves data quality in many cases, in terms of classification accuracy, over existing label noise correction methods.**

Keywords—Label noise，Internet of Things, Medical auxiliary diagnosis system，ensemble learning

I. INTRODUCTION

Machine learning (ML) and the Internet of Things (IoT) have been changing the way of our life from different aspects[1-5]. Adopting these technologies in healthcare can bring many opportunities for medical IT. Nowadays, many hospitals combine machine learning and IoT technology to obtain data which was used into hospital information management system (HIS)[6]. Hospitals store a large amount of medical data which are recorded during the diagnosis and treatment of patients in the database of HIS system. Faced with such a large and complicated medical data set, how to construct a high-quality medical diagnosis system driven by data classification model has become a hot topic.

It is generally accepted that the highest accuracy HIS can achieve depends on the quality of the data and the learning algorithm. As shown by Malossini et al.[7], even a very small number of mislabeled samples can seriously degrade the performance of the obtained classifier. In addition, the smaller the sample size, the greater the impact of mislabeled samples. Therefore, in this paper, we focus on the label noise, that is, the influence of a certain number of wrong mark samples in the data set on the performance of supervised learning.

First of all, in addition to data input errors, when the amount of information provided to experts is small and the quality is poor, experts may also make wrong judgments.[8, 9]. Moreover, the amount of available information will be reduced. when the description language is restricted[10].Secondly, as long as human experts are involved, label noise will naturally occur[11]. Medical experts may make different subjective judgments regarding unquantifiable pathological features[12-14].

In general, the description of cases is characterized by strong subjectivity and insufficient information, which constitute the main influencing factors for the false label in the dataset. Edmonds[15] mentioned that noise is usually a complex phenomenon, but there is no completely specified label noise model, which indicates that we need to find an automated algorithm that can deal with label noise.

In response to this problem, machine learning researchers have conducted two studies, (1) Algorithm-level research (2) Data-level research. Compared to highly robust algorithms, the label filtering algorithm has the characteristics of convenience and simplicity, and could achieve good accuracy. However, for the dataset with high noise, the label filtering algorithm could greatly reduce the number of samples, leading to the aggravation of the imbalance of samples, the reduction of rare samples and other new problems. Therefore, this paper proposes a hybrid label noise correction algorithm, which can withstand the negative effects of random noise and non-random label noise. We adopt the idea of ensemble learning[16-18] and design a label correction algorithm by combining two perspectives of clustering and classification. A large amount of literature proves that ensemble learning usually achieves better results compared to a single model. Then we re-label the labeled noise samples, and finally use all the non-labeled noise samples to train the supervise learning algorithm to complete the classification task. In general, our contributions can be highlighted as follows.

(1) We propose a hybrid label noise correction algorithm to deal with partial label noise. We use Pseudo labels to make the most of the available data. By ensembling classification semi-supervision and clustering semi-supervision, the label correction of noise samples is carried out from two perspectives, which can significantly improve the accuracy of noise label relabeling.

(2) Considering that data label noise often appears in the dataset used to train the auxiliary medical diagnosis system, and the label noise will seriously affect the performance of the system. We proposed Medical auxiliary diagnosis system with our approach. The experimental results show that this method has better robustness and accuracy than the existing algorithms.

The rest of this paper is organized as follows, section II introduced the processing method for the label noise dataset. Section III introduced the proposed hybrid label noise correction algorithm in detail and described it mathematically, then introduced a Medical auxiliary diagnosis model based on our method. Section IV report and analyze experimental results on benchmark and real-world medical datasets. Finally, the conclusion is drawn in section V.

II. RELATED WORK

Methods to deal with label noise have been widely studied. This section will briefly review the treatment methods of label noise, more detailed introduction can refer to the recent literature review[19]. There are two ways to deal with label noise.

A. Algorithm-level approaches

Algorithm-level approaches can be specially modified to reduce the effect of label noise on classifier and tolerate the presence of noise. All these algorithms have to label the noise data has the characteristics of the inherently robustness. For example, the classic AdaBoost usually gives large weights to samples with error markers, but a simple tolerance extension corrects the size of the instance weight. Therefore, for the label noise sensitivity of AdaBoost algorithm, many related robust algorithms are proposed. Domingo and Watanabe[20] proposed MadaBoost algorithm in 2000. In 2008, Bradley and Schapire[21] proposed the Filter Boost algorithm, which adopted logarithmic loss function instead of exponential loss function. In recent years, Rusiecki[22] proposed a deep convolutional neural network for classified cross-entropy loss function pruning training method. Aiming at the problem of label noise in medical images, Cheng Xue et al.[23] designed a sample reweighting strategy to maintain the usefulness of correctly labeled hard samples. The method has been proved to be effective in the classification of skin lesions. Unfortunately, when noise rates are high, many robust algorithms are not available and their performance is poor. These methods are based on theoretical methods and allow us to use the knowledge gained by analyzing the consequences of tag noise, but at the cost of increasing the complexity of the learning algorithm. The data level approach is a better choice, and cleansing data can be applied in many situations.

B. Data-level approach

Data-level approaches, which aims to filter or clean up noise samples before the training classifier[9, 24]. As shown in literature[24, 25], noise filters include identifying and deleting instances of marked errors in training Data. In the literature, removing noise labels is more common than correcting them[26].It may seem easier to do this by simply removing error samples, but in some cases, noise sample filtering methods can remove too much data. This could degrade classification performance[26] . In addition, some data are expensive and difficult to collect, so only a small amount of tagged data is provided[27] .Therefore, in this case, it is more appropriate to correct the label noise. Compared with noise filter, noise correction is not to delete the noise sample, but to correct it. However, there are few studies on the correction of label noise. Literature[28, 29] proposed two algorithms to solve this problem. Literature[30] proposes a novel algorithm called " adaptive voting noise correction (AVNC) ', which is used to accurately identify and correct noise labels, but it is only applicable to binary classification problems. Literature[31] uses a joint optimization framework that finds and corrects noise labels. However, this way still requires a large amount of data as a prerequisite, resulting in relatively poor performance when using small datasets. At the same time, modifying the label is not satisfactory. Inspired by the recursive ensemble learning approach, we intend to build a general algorithm for correcting noise labels from two perspectives, clustering semi-supervised and classification semi-supervised, which can handle not only binary classification problems but also multi-classification problems.

III. DESCRIPTION OF OUR APPROACH

In this section, we mainly describe the details of hybrid label noise correction algorithm (HLNC) and Medical auxiliary diagnosis system based on our algorithm.

A. hybrid label noise correction algorithm

In the proposed strategy, dividing high-confidence data (real label samples) and low-confidence data (noise samples) from the noisy dataset is an important part. We employ the k-means algorithm to divide the dataset into K clusters based on the Euclidean distance and square error sum. The center point C_j of each cluster can be expressed as Eq (1).

$$C_j = \frac{1}{N_j}\sum_{i=1}^{N_j} x_i \text{ }, j = 1, \dots, K \tag{1}$$

Where N_j represents the sample size in the j-th cluster. So, the objective function of clustering is Eq (2).

$$J = \sum_{j=1}^{K}\sum_{i=1}^{N_k} d\left(x_i, C_j\right)^2 \tag{2}$$

Where $d((x_i, C_j)$ is the Euclidean distance between x_i and C_j. Our next work is to determine the label m_j of each cluster. Next, compare each sample in the cluster with the cluster label, the same is a high-confidence sample, and the different is a low-confidence sample.

We use high-confidence samples to train the SSK-means and Co-training algorithms, while erasing the labels of low-confidence samples. Basu et al[32] introduced the seed set $S = \cup_{h=1}^{K} S_h$ formed by a small number of labeled samples, and

Identify applicable funding agency here. If none, delete this text box.

divided the seed set into K clusters, and then initialized the K-means algorithm to form the SSK-means semi-supervised algorithm. First use label data to get K cluster centers.

$$c_h = \frac{1}{|S_h|} \sum_{x \in S_h} x \, , h = 1, \ldots, K \tag{3}$$

Then assign unlabeled data for clustering

$$h^* = arg_h \min\{d(x, c_h)^2\} \tag{4}$$

Finally, use the Eq (5) to recalculate the cluster center, and iterate until convergence. While training SSK-means, we also train the co-training algorithm. Blum et al[33] first proposed the Co-training algorithm, which uses two classifiers for collaborative training. Therefore, we choose two heterogeneous algorithms including decision tree (DT) and Support Vector Machine (SVM) as our basic classifiers. The loss function of DT and SVM is recorded as Eq. (6) ~ (7).

$$L_i(DT) = \sum_{t=1}^{T} N_t H_t(T) + \alpha |T| \tag{5}$$

$$L_i(SVM) = \sum_{j \neq y_i} \max\left(0, w_j^T x_i - w_{y_j}^T x_i + b\right) \tag{6}$$

Where, N_t is the number of leaf node samples， $H_t(T)$ is the entropy value on leaf node t, and $\alpha \geq 0$ is the parameter， w_j^T stand for the input, $f_{y_i} = w_{y_i}^T x$, $f_j = w_j^T x$, b is the super parameter of SVM and y is the prediction output.

For the Co-training algorithm, maintaining the difference between the two classifiers is an important factor determining the effectiveness of the algorithm, so the feature subset used to train the two classifiers should satisfy eq（8）

$$P(x_1 | f(x), x_2) = P\left(x_1 | f(x)\right) \tag{7}$$

Where P (*) represents probability. When two feature subsets satisfy conditional independence, the co-training algorithm can use unlabeled samples to improve the accuracy of the classifier.

Finally, we compare the labels obtained based on the SSK-means algorithm and the Co-training algorithm. If the results are the same, the sample will be labeled as a high-confidence sample; if they are different, the sample will be directly put back into the low-confidence data. Then iterate the label correction process until there are no low confidence samples or stop after a certain number of times. Finally output the high confidence samples to complete the label correction task. For different datasets, a classification algorithm more suitable for data can be used for testing. Because the random forest algorithm has good generalization ability, we choose the random forest algorithm for experiments. We show the flow chart of the HLNC algorithm in Figure 1.

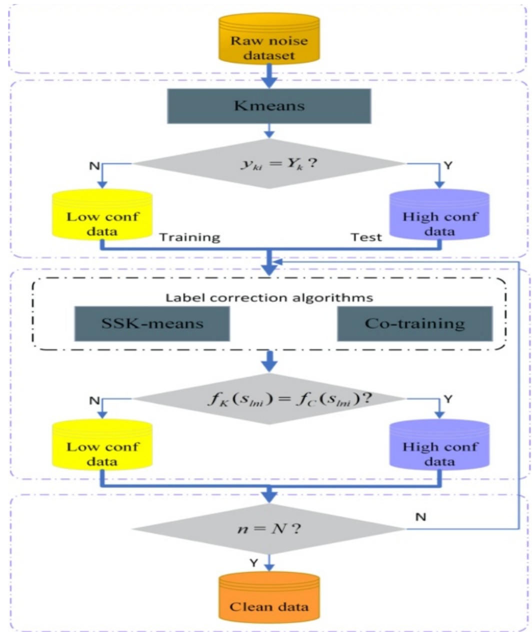

Fig. 1 Overview of the hybrid label noise correction algorithm(HLNC)

B. *Medical auxiliary diagnosis system with HLNC*

A large amount of medical data is stored in the HIS system, which establishes a strong foundation for data-driven medical auxiliary diagnostic system. However, because the label noise problem often occurs in medical data, and the label noise will seriously affect the performance of the medical assistant diagnostic system, we have designed a medical assistant diagnostic system based on HLNC.

As shown in Figure 2, the system simulates a traditional diagnostic process and has four modules. including data acquisition, data preprocessing, algorithm diagnosis and result output. The main purpose of this system is to solve the problem of medical data label noise and effectively improve the robustness and accuracy of medical diagnostic systems. Using this system can not only improve the accuracy and efficiency of medical diagnosis, but also alleviate the shortage of medical resources.

IV. EXPERIMENTAL EVALUATION

To validate the method presented in section III, we conducted extensive evaluation and experiments with a variety of label noise tasks, and report our results in this section. The extensive experiments are designed to include three phases, (1) with the noise sensitive classifier as the object, the results of using our method and not using our method are compared under the condition of different noise rates, and the effectiveness of our method is verified; (2) Compared with other label noise correction algorithms, our method has better classification performance; (3) further experiments are carried out to facilitate comparative testing with two compared algorithms on real-world dataset (cerebral stroke datasets).

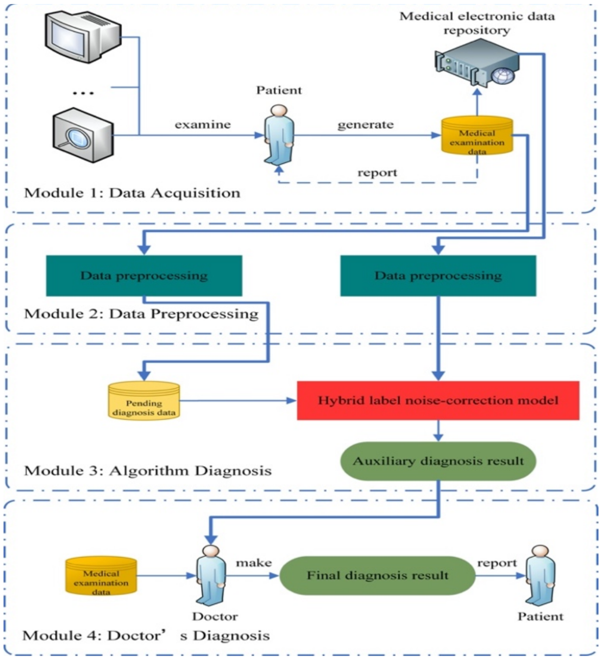

Fig. 2 Medical auxiliary diagnosis system with HLNC

In order to verify the label noise in our problem, we designed a label noise generation method based on data sample proportion. Different from simply generating random noise, our method will increase the noise according to the proportion relation of the data set samples, so as to better detect the anti-noise performance of the algorithm. In this paper, the noise rate is set to no more than 50%, if more than that, supervised learning cannot be carried out[26].

In our experiment, we used 8 medical standard benchmark datasets from the UCI repository[34] and a real-world dataset for validation. The 10-fold cross validation method was used for the evaluation, and the average value of the accuracy obtained indicated the generalization ability of the evaluated classifier in the test.

TABLE I. BENCHMARK DATASETS INFORMATION

Dataset	Attribute	Instance
Adolescent-Autism	20	97
Mammographic masses	5	829
breast-cancer	9	276
Cryotherapy	6	89
Wdbc	30	568
breast-cancer-wisconsin	9	682
Heart	13	269
SpectHeart	22	266

A. *Algorithm Effectiveness Evaluation*

In order to prove that label noise has a serious impact on the classifier, and our method can well help the classifier to conduct training in the environment with label noise, we increased the noise rate of 10%,20%,30%,40% and 50% respectively on the eight benchmark datasets.

We use two noise sensitive classifier decision trees (DT) and linear discriminant analysis (LDA) algorithm, and the random forest (RF) algorithm with some noise resistance is used as the basic classifier to directly apply to the noise datasets. Then, the high-quality dataset processed by HLNC algorithm was applied to the above three basic algorithms, and the three classifiers were retrained under the same parameter environment. Finally, we compared the accuracy of the two groups of classifiers by the 10-fold cross validation method.

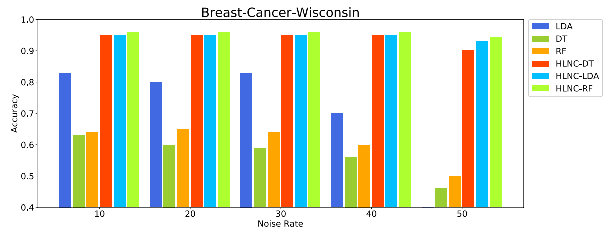

Fig. 3. Accuracy of different noise rate on the Breast-Cancer-Wisconsin datasets

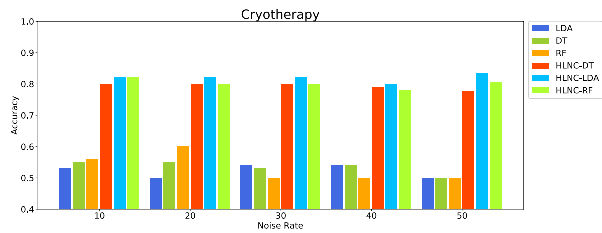

Fig. 4. Accuracy of different noise rate on the Cryotherapy datasets

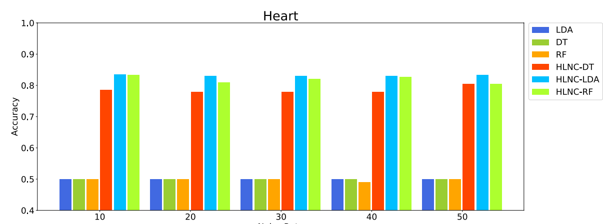

Fig. 5. Accuracy of different noise rate on the Heart dataset

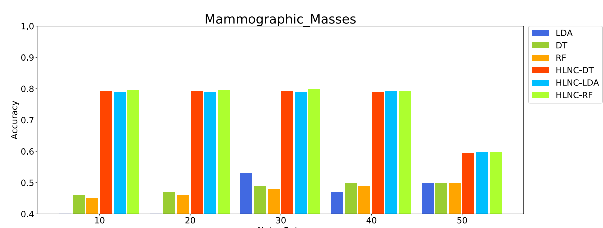

Fig. 6. Accuracy of different noise rate on the Mammographic masses dataset

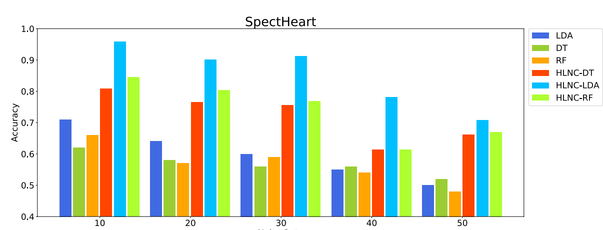

Fig. 7. Accuracy of different noise rate on the SpectHeart dataset

As shown in figure 3-10, when the three basic classifiers are directly applied to the label noise datasets, the performance of the classifier will significantly decrease with the increase of noise rate. In some tests, the basic classifier accuracy rate is less than 50% and has lost the ability to classify, so it can be concluded that the influence of label noise on the classifier is very serious. However, through our algorithm, a large number of label noise has been effectively corrected. The results show that our algorithm is effective in correcting label noise.

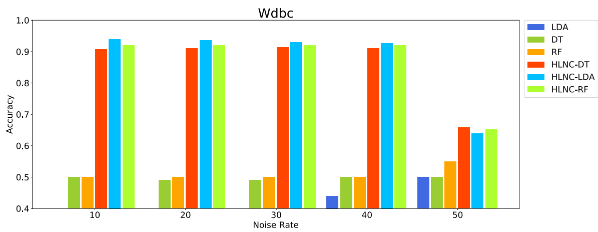

Fig. 8. Accuracy of different noise rate on the Wdbc dataset

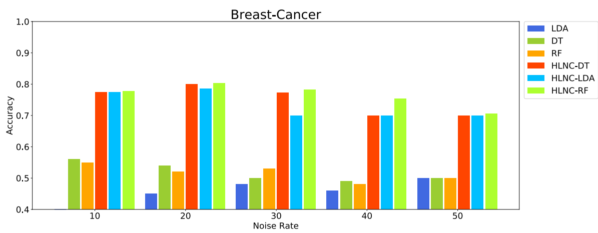

Fig. 9. Accuracy of different noise rate on the Breast-cancer dataset

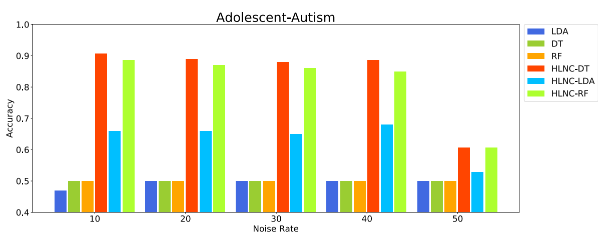

Fig. 10. Accuracy of different noise rate on the Adolescent-Autism dataset

B. *Algorithm performance evaluation*

In order to further prove the performance of our method, we chose the novel label noise correction algorithm CL algorithm[35] and RMDA algorithm[27] for comparison on the above eight medical datasets. For CL and HLNC algorithms, we chose random forest as the basic classifier, where the random forest algorithm use default parameters.

As shown in table 1, the classification accuracy of our method is generally better than LNL and RMDA, which means that our method has certain advantages in performance. The results show that our algorithm can effectively correct noise label, improve data quality and have strong generalization ability.

C. *Experiments with cerebral stroke datacase*

In this part, we applied our method to the stroke database to illustrate its practical potential in medical auxiliary diagnosis system. The stroke database was obtained from a hospital in Kunming, China, using an RFID device to test subjects for one year. The experiments were conducted in accordance with the highest ethical standards, both locally (bioethics standards provided by Yunnan provincial people's hospital) and internationally (documents provided by the world health organization), and were always supervised by expert physicians. The dataset records the data of 3438 stroke patients, which includes the patient's basic information (such as birth date, age, occupation, etc.), stroke history, heart disease history and type, diabetes, hypertension, smoking and drinking, physical exercise Wait, a total of 199 dimensions. After a series of operations such as data encoding, data cleaning, and data integration, we selected 33 highly correlated attributes to form the final training data.

TABLEII. ACCURACY OF DIFFERENT ALGORITHMS ON THE EIGHT MEDICAL DATASETS

Dataset	Algorithm	Noise rate				
		10%	*20%*	*30%*	*40%*	*50%*
Adolescent-Autism	CL	0.526	0.526	0.474	0.526	0.474
	RMDA	0.842	0.842	0.789	0.789	0.526
	HLNC	**0.887**	**0.87**	**0.86**	**0.85**	**0.607**
Mammographic masses	CL	0.497	0.526	0.679	0.739	**0.788**
	RMDA	0.327	0.6	0.673	0.745	0.745
	HLNC	**0.795**	**0.795**	**0.8**	**0.794**	0.598
Breast-cancer	CL	**0.782**	0.745	0.673	0.636	0.636
	RMDA	0.8	0.727	0.727	0.527	0.612
	HLNC	0.777	**0.803**	**0.782**	**0.753**	**0.705**
Cryotherapy	CL	**0.882**	**0.824**	0.765	0.765	0.706
	RMDA	0.824	0.765	0.765	0.765	0.647
	HLNC	0.821	0.8	**0.8**	**0.78**	**0.806**
Wdbc	CL	0.504	0.549	0.566	0.611	0.655
	RMDA	0.319	0.327	0.327	0.699	**0.673**
	HLNC	**0.92**	**0.92**	**0.92**	**0.92**	0.652
Breast-cancer-Wisconsin	CL	0.897	0.897	0.949	0.934	0.919
	RMDA	0.897	0.897	0.897	0.897	0.846
	HLNC	**0.96**	**0.96**	**0.96**	**0.96**	**0.942**
Heart	CL	0.753	0.683	0.642	0.52	0.53
	RMDA	0.63	0.61	0.56	0.54	0.57
	HLNC	**0.845**	**0.803**	**0.769**	**0.613**	**0.67**
SpectHeart	CL	0.753	0.683	0.642	0.52	0.53
	RMDA	0.63	0.61	0.56	0.54	0.57
	HLNC	**0.845**	**0.803**	**0.769**	**0.613**	**0.67**

In order to evaluate the classification performance of stroke datasets, we performed missing value processing, normalized and abnormal data removal. Then, according to our label noise generation method, add noise to it. Our method is compared with the previous label noise correction algorithm. As shown in table 2, HLNC still achieves the best performance among all the compared approaches for real-world medical dataset.

TABLE III. ACCURACY OF DIFFERENT ALGORITHMS ON THE STROKE DATABASE

Algorithm	Noise rate				
	10%	20%	30%	40%	50%
CL	**0.965**	0.958	0.916	0.866	0.654
RMDA	0.722	0.719	0.722	0.69	0.719
HLNC	0.961	**0.96**	**0.96**	**0.954**	**0.952**

V. CONCLUSION

In this paper, we propose a hybrid label noise correction algorithm to support IoT-based medical diagnosis system. Its main design purpose is to solve the problem of label noise in most data collected from IoT-based system. As shown in the experiments, when label noise appears in datasets, especially in medical data, our method can effectively improve data quality and its performance is better than the latest algorithm. As a result, it provides a promising and easy-to-use technology for the system, which can effectively solve the following problems, due to the subjectivity and lack of information in the description of cases, label noise appears in the medical data.

First, we propose a hybrid label noise correction algorithm. By ensembling the results of clustering semi-supervision and classification semi-supervision, we can better correct the samples marked wrong. Finally, through the experimental results of the medical datasets, it is proved that HLNC algorithm provides a potential solution for the data analysis part of the medical auxiliary diagnosis system and has application value.

Although the experimental results show that our method has achieved a certain degree of superiority in medical datasets, there are still some problems that may lead to the practical effect. Similar to all existing noise label correction algorithms, the proposed method may have poor performance when the noise ratio is large and the sample imbalance in the dataset is serious. In the future, we will explore the complementarity of different learning algorithms to make the best choice for basic learners.

References

[1] A. Darwish, A. E. Hassanien, M. Elhoseny, A. K. Sangaiah, and K. Muhammad, "The impact of the hybrid platform of internet of things and cloud computing on healthcare systems: opportunities, challenges, and open problems," Journal of Ambient Intelligence & Humanized Computing, 2017.

[2] K. Muhammad, R. Hamza, J. Ahmad, J. Lloret, H. H. G. Wang, and S. W. Baik, "Secure Surveillance Framework for IoT systems using Probabilistic Image Encryption," IEEE Transactions on Industrial Informatics, pp. 1-1, 2018.

[3] J. Qi, P. Yang, L. Newcombe, X. Peng, Y. Yang, and Z. Zhao, "An Overview of Data Fusion Techniques for Internet of Things enabled Physical Activity Recognition and Measure," Information Fusion, vol. 55, pp. 269-280, 2020.

[4] J. Qi, P. Yang, A. Waraich, Z. Deng, Y. Zhao, and Y. Yang, "Examining sensor-based physical activity recognition and monitoring for healthcare using Internet of Things: A systematic review," Journal of Biomedical Informatics, vol. 87, pp. 138-153, 2018.

[5] J. Qi, P. Yang, M. Hanneghan, S. Tang, and B. Zhou, "A Hybrid Hierarchical Framework for Gym Physical Activity Recognition and Measurement Using Wearable Sensors," IEEE Internet of Things Journal, vol. 6, no. 2, pp. 1384-1393, 2019.

[6] G. White, R. Hill, M. J. Zakrewski, R. Kummerlen, M. S. Abbott, and R. C. Brooks, "Patient medication IV delivery pump with wireless communication to a hospital information management system," ed: Google Patents, 2004.

[7] A. Malossini, E. Blanzieri, and R. Ng, "Assessment of SVM reliability for microarray data analysis," University of Trento, 2004.

[8] R. J. Hickey, "Noise modelling and evaluating learning from examples," Artificial Intelligence, vol. 82, no. 1-2, pp. 157-179, 1996.

[9] C. E. Brodley and M. A. Friedl, "Identifying mislabeled training data," Journal of artificial intelligence research, vol. 11, pp. 131-167, 1999.

[10] P. Brazdil and P. Clark, "Learning from imperfect data," in Machine Learning, Meta-Reasoning and Logics: Springer, 1990, pp. 207-232.

[11] D. McNicol, "What are statistical decisions," in A Primer of Signal Detection Theory: Allen, 1972, pp. 1-17.

[12] A. Malossini, E. Blanzieri, and R. T. Ng, "Detecting potential labeling errors in microarrays by data perturbation," Bioinformatics, vol. 22, no. 17, pp. 2114-2121, 2006.

[13] N. P. Hughes, S. J. Roberts, and L. Tarassenko, "Semi-supervised learning of probabilistic models for ECG segmentation," in The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2004, vol. 1: IEEE, pp. 434-437.

[14] P. Yang, J. Liu, J. Qi, Y. Yang, X. Wang, and Z. Lv, "Comparison and Modelling of Country-level Microblog User and Activity in Cyber-physical-social Systems Using Weibo and Twitter Data," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 10, no. 6, pp. 1-24, 2019.

[15] F. Squazzoni, Epistemological aspects of computer simulation in the social sciences. Springer, 2009.

[16] Y. Yang and J. Jiang, "Adaptive Bi-weighting toward automatic initialization and model selection for HMM-based hybrid meta-clustering ensembles," IEEE transactions on cybernetics, vol. 49, no. 5, pp. 1657-1668, 2018.

[17] Y. Yang, X. Liu, Q. Ye, and D. Tao, "Ensemble Learning-Based Person Re-identification with Multiple Feature Representations," Complexity, vol. 2018, 2018.

[18] Y. Yang and X. Liu, "A robust semi-supervised learning approach via mixture of label information," Pattern Recognition Letters, vol. 68, pp. 15-21, 2015.

[19] R. C. Prati, J. Luengo, and F. Herrera, "Emerging topics and challenges of learning from noisy data in nonstandard classification: a survey beyond binary class noise," Knowledge and Information Systems, vol. 60, no. 1, pp. 63-97, 2019.

[20] C. Domingo and O. Watanabe, "MadaBoost: A modification of AdaBoost," in COLT, 2000, pp. 180-189.

[21] J. K. Bradley and R. E. Schapire, "Filterboost: Regression and classification on large datasets," in Advances in neural information processing systems, 2008, pp. 185-192.

[22] A. Rusiecki, "Trimmed categorical cross-entropy for deep learning with label noise," Electronics Letters, 2019.

[23] C. Xue, Q. Dou, X. Shi, H. Chen, and P. A. Heng, "Robust Learning at Noisy Labeled Medical Images: Applied to Skin Lesion Classification," arXiv preprint arXiv:1901.07759, 2019.

[24] D. Gamberger, N. Lavrac, and C. Groselj, "Experiments with noise filtering in a medical domain," in ICML, 1999, pp. 143-151.

[25] D. Guan, W. Yuan, Y.-K. Lee, and S. Lee, "Identifying mislabeled training data with the aid of unlabeled data," Applied Intelligence, vol. 35, no. 3, pp. 345-358, 2011.

[26] B. Frénay and M. Verleysen, "Classification in the presence of label noise: a survey," IEEE transactions on neural networks and learning systems, vol. 25, no. 5, pp. 845-869, 2014.

[27] C. Bouveyron and S. Girard, "Robust supervised classification with mixture models: Learning from data with uncertain labels," Pattern Recognition, vol. 42, no. 11, pp. 2649-2658, 2009.

[28] B. Nicholson, J. Zhang, V. S. Sheng, and Z. Wang, "Label noise correction methods," in 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2015: IEEE, pp. 1-9.

[29] B. Nicholson, V. S. Sheng, and J. Zhang, "Label noise correction and application in crowdsourcing," Expert Systems with Applications, vol. 66, pp. 149-162, 2016.

[30] J. Zhang, V. S. Sheng, T. Li, and X. Wu, "Improving crowdsourced label quality using noise correction," IEEE transactions on neural networks and learning systems, vol. 29, no. 5, pp. 1675-1688, 2017.

[31] D. Tanaka, D. Ikami, T. Yamasaki, and K. Aizawa, "Joint optimization framework for learning with noisy labels," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5552-5560.

[32] S. Basu, A. Banerjee, and R. Mooney, "Semi-supervised clustering by seeding," in In Proceedings of 19th International Conference on Machine Learning (ICML-2002, 2002: Citeseer.

[33] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in Proceedings of the eleventh annual conference on Computational learning theory, 1998, pp. 92-100.

[34] A. Asuncion and D. Newman, "UCI machine learning repository," ed, 2007.

[35] C. G. Northcutt, L. Jiang, and I. L. Chuang, "Confident Learning: Estimating Uncertainty in Dataset Labels," arXiv preprint arXiv:1911.00068, 2019.