A Pathological Diagnosis Method for FUO based on Multi-Path Hierarchical Classification: Model Design and Validation

Jianchao Du, Junyao Ding, Yuan Wu, Tianyan Chen, Jianqi Lian, Lei Shi, Yun Zhou

Submitted to: JMIR Formative Research
on: April 30, 2024

Table of Contents

A Pathological Diagnosis Method for FUO based on Multi-Path Hierarchical Classification: Model Design and Validation

Jianchao Du[1]; Junyao Ding[1]; Yuan Wu[2]; Tianyan Chen[3]; Jianqi Lian[4]; Lei Shi[3]; Yun Zhou[4]

[1]School of Telecommunications Engineering University of Xidian Xi'an CN
[2]Duke University Health System Durham US
[3]Department of infectious diseases The First Affiliated Hospital of Xi'an Jiaotong University Xi'an CN
[4]Department of infectious diseases The Second Affiliated Hospital of Air Force Medical University Xi'an CN

Corresponding Author:
Yun Zhou
Department of infectious diseases
The Second Affiliated Hospital of Air Force Medical University
Xinsi Road
Xi'an
CN

Abstract

Background: Fever of Unknown Origin (FUO) is a significant challenge for the medical community due to its association with a wide range of diseases, the complexity of diagnosis, and the likelihood of misdiagnosis. Machine learning can extract valuable information from the extensive data of patient indicators, aiding doctors in diagnosing the underlying cause of FUO.

Objective: A hierarchical classification method based on multipath feature selection was proposed to address the problem that the causes of FUO are diverse and difficult to diagnose accurately.

Methods: The dataset was obtained from the clinical diagnostic records of patients with FUO who were admitted to the First Affiliated Hospital of Xi'an Jiaotong University between the years 2011 and 2020. The dataset contains 564 samples, with 5 broad categories of etiologies and subcategories, totaling 16 precise etiologies. An intelligent diagnostic algorithm for FUO was constructed based on hierarchical classification to identify 16 precise etiologies and two improvements were made: (1) A multi-path prediction model to minimize the probability of error propagation; (2) A utilization of L1,2 regularization constraint to effectively identify the most suitable subset of features to eliminate redundancy and interference in the hierarchical classification process. To confirm the proposed method's validity, we conducted ablation experiments and compared it with other flat and hierarchical classification algorithms.

Results: According to the ablation experiments, the proposed method achieves the optimal result when the intermediate path is 3, with an accuracy of 72.35%, FH of 85.01%, and FLCA of 83.29%. This performance is superior to the traditional single-path hierarchical classification, showing improvements of 5.69%, 2.89%, and 3.39%, respectively. The feature selection based on the L1,2 regularization constraint can further enhance the model's performance. The best results were achieved by filtering 25% of the features, with an accuracy of 76.08%, FH of 86.72%, and FLCA of 85.39%. Based on the comparison experiments, the proposed method outperforms the seven flat algorithms and four hierarchical classification algorithms that were compared. For example, the accuracy can be improved by at least 5.39% compared to ELM and up to 24.7% compared to KNN.

Conclusions: The proposed algorithm greatly improves the classification performance compared to existing machine learning algorithms. It can better predict the cause of FUO, assisting physicians in their work.

(JMIR Preprints 30/04/2024:58423)
DOI: https://doi.org/10.2196/preprints.58423

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users. Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

111 1 1A Pathological Diagnosis Method for FUO based on Multi-Path Hierarchical Classification: Model Design and Validation

Abstract

Background: Fever of Unknown Origin (FUO) is a significant challenge for the medical community due to its association with a wide range of diseases, the complexity of diagnosis, and the likelihood of misdiagnosis. Machine learning can extract valuable information from the extensive data of patient indicators, aiding doctors in diagnosing the underlying cause of FUO.
Objective: A hierarchical classification method based on multipath feature selection was proposed to address the problem that the causes of FUO are diverse and difficult to diagnose accurately.
Methods: The dataset was obtained from the clinical diagnostic records of patients with FUO who were admitted to the First Affiliated Hospital of Xi'an Jiaotong University between the years 2011 and 2020. The dataset contains 564 samples, with 5 broad categories of etiologies and subcategories, totaling 16 precise etiologies. An intelligent diagnostic algorithm for FUO was constructed based on hierarchical classification to identify 16 precise etiologies and two improvements were made: (1) A multi-path prediction model to minimize the probability of error propagation; (2) A utilization of $L_{1,2}$ regularization constraint to effectively identify the most suitable subset of features to eliminate redundancy and interference in the hierarchical classification process. To confirm the proposed method's validity, we conducted ablation experiments and compared it with other flat and hierarchical classification algorithms.
Results: According to the ablation experiments, the proposed method achieves the optimal result when the intermediate path is 3, with an accuracy of 72.35%, F_H of 85.01%, and F_{LCA} of 83.29%. This performance is superior to the traditional single-path hierarchical classification, showing improvements of 5.69%, 2.89%, and 3.39%, respectively. The feature selection based on the $L_{1,2}$ regularization constraint can further enhance the model's performance. The best results were achieved by filtering 25% of the features, with an accuracy of 76.08%, F_H of 86.72%, and F_{LCA} of 85.39%. Based on the comparison experiments, the proposed method outperforms the seven flat algorithms and four hierarchical classification algorithms that were compared. For example, the accuracy can be improved by at least 5.39% compared to ELM and up to 24.7% compared to KNN.
Conclusions: The proposed algorithm greatly improves the classification performance compared to existing machine learning algorithms. It can better predict the cause of FUO, assisting physicians in their work.

Keywords: Fever of Unknown Origin (FUO); Intelligent Diagnosis; Machine Learning; Hierarchical Classification; Feature Selection.

Introduction

Background

Fever of Unknown Origin[1] (FUO) is a medical term utilized to describe a group of diseases that exhibit a prolonged fever lasting for a duration exceeding three weeks, and which cannot be diagnosed even after one week of outpatient or inpatient examinations. This concept was initially introduced by Petersdorf and Beeson in 1961. The etiology of FUO is multifactorial and encompasses a wide range of factors, including over 200 different species[2], such as streptococcus pneumoniae[3], peritoneal mesothelioma[4], bacteroides fragilis[5]. The distribution of these causative agents varies both temporally and geographically, necessitating comprehensive and in-depth investigations to accurately determine the underlying cause of the

disease. Consequently, identifying the cause of FUO poses a significant challenge within the medical field[6]. In the diagnosis of febrile illness, doctors need to conduct a thorough evaluation and examination based on the patient's symptoms, signs and possible causes to determine the final diagnosis and treatment plan. However, despite conducting a comprehensive examination, it has been found that one-thirds of patients presenting with persistent fever remain undiagnosed[7].

With the continuous progress of machine learning (ML), its application in various domains of production and business activities has experienced a substantial growth[8-10]. In the realm of medicine, the utilization of ML-based disease diagnosis technology holds immense importance as it aids in enhancing the accuracy and real-time capabilities of doctors' diagnoses. In recent years, there has been a significant increase in the advancement of intelligent diagnostic techniques that employ ML algorithms to independently predict potential causes of diseases. Reference [11] employed a logistic regression model for the purpose of classifying and diagnosing cases of malignant pleural mesothelioma. Reference [12] introduced the utilization of the grey wolf algorithm for the purpose of feature selection, aiming to eliminate redundant and irrelevant features within the dataset. Additionally, the authors employed SVM to classify the data related to coronary artery disease. Reference [13] proposed a distance-based KNN approach for the classification of heart disease data. The authors recommended the utilization of a distance set that incorporates multiple distance calculations, in addition to a weighted voting mechanism, to enhance the predictive accuracy of conventional KNN algorithms. Reference [14] proposed a liver disease classification approach that utilizes the XGBoost algorithm, and further improved the performance of the XGBoost model by optimizing its parameters using a genetic algorithm. Reference [15] introduced a classification method for Parkinson's disease that utilizes the synthetic minority oversampling technique. This technique aims to address the issue of imbalanced class distribution by transforming it into a balanced distribution. Additionally, the dataset of Parkinson's disease was utilized for binary classification through the implementation of the random forest algorithm, resulting in enhanced accuracy in classifying the minority class and overall improvement in classification performance.

The advantage of artificial intelligence resides in its capacity to continuously amass clinical case data and derive insights from it, ultimately constructing robust diagnostic models. Therefore, the development of an intelligent diagnostic model for FUO would significantly mitigate the challenges related to the diagnosis and treatment of this condition.

Objectives

Due to the current major applications being primarily binary classification or a few-class classification problems, algorithm designs are often directly aimed at all classes using flat classification methods. However, FUO can be attributed to numerous potential causes, thereby necessitating a multi-class classification approach. Consequently, employing flat classification methods alone results in suboptimal accuracy and fails to fulfill the application's requirements[16]. Given the hierarchical structure of the etiological labels in the FUO dataset, it is possible to employ hierarchical classification[17] models for the analysis of the dataset. By leveraging the hierarchical associations among data classes, a top-down methodology is employed for hierarchical classification, culminating in the acquisition of the corresponding class at the leaf level. Hierarchical classification involves the decomposition of a multi-class task into several sub-classification tasks, resulting in a simplified model and reduced complexity in modeling. Additionally, it presents a notable advantage in terms of computational

efficiency for both classification learning and prediction tasks. This characteristic makes it particularly suitable for fulfilling the requirements of etiological prediction in cases of FUO.

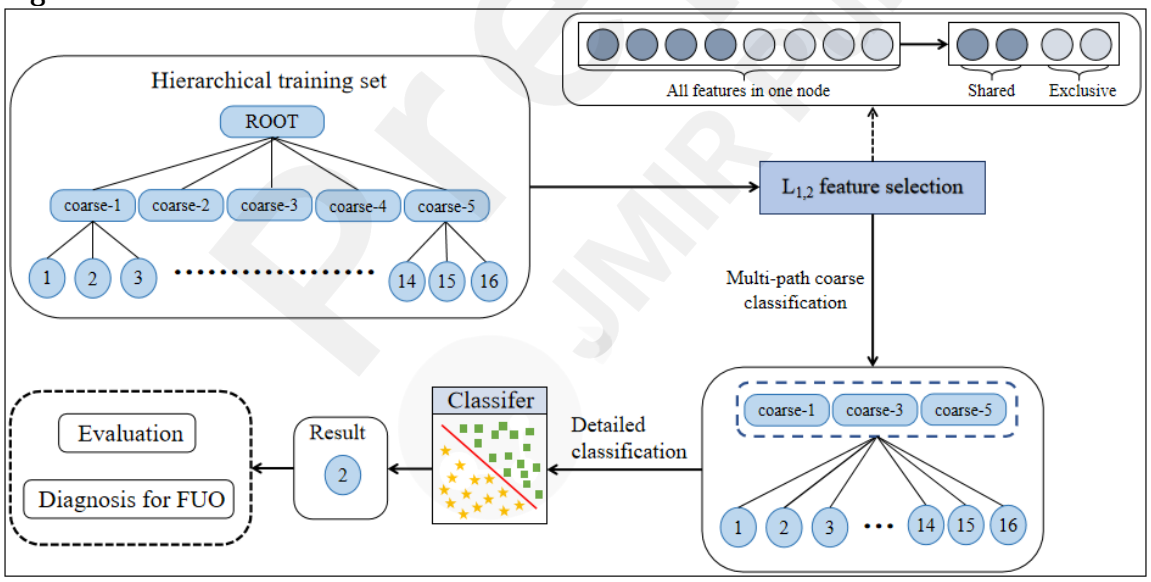
Therefore, we introduce a novel auxiliary diagnostic method for FUO utilizing multi-path feature selection and hierarchical classification. The data will be organized into a hierarchical structure based on disease classes for the purpose of hierarchical classification. Subsequently, prediction will be conducted from the highest level to the lowest level until the final classification class is achieved. To mitigate the likelihood of ineffective lower-level classification resulting from errors in higher-level classification, the hierarchical classification process incorporates multiple path prediction models with controllable pre-selected classes. This approach aims to enhance the accuracy of lower-level classification. Additionally, $L_{1,2}$ regularization constraint[18] is utilized for the purpose of feature selection at each level in order to eliminate redundant features and subsequently minimize interference, thereby enhancing the accuracy of prediction.

Methods

Framework

The framework of the hierarchical classification method based on multi-path and feature selection (MPFSHC) proposed in this paper is illustrated in Figure 1. The process can be divided into two steps: (1) Feature selection is performed at each layer using $L_{1,2}$ regularization constraints based on the tree hierarchy to eliminate redundant features and reduce interference. Moreover, we innovatively select layer-shared features and class-exclusive features for each node; (2) Hierarchical classification is then performed using the selected features, and multi-path prediction models are built by pre-selecting controllable multiple classes during the hierarchical classification process.

Figure 1. Framework of MPFSHC



Hierarchical feature selection

Traditional feature selection is based on the assumption that all classes are independent of each other, and a set of common features is selected for all classes to form a subset of features before classification. However, reference [19] found that certain features are more suitable for the classification of some classes with better discriminative properties. On the other hand, these features do not improve the classification performance for other classes. Feature selection in hierarchical structures allows for the selection of a distinct subset of features for

each sub-categorization task within the structure. This approach enhances the performance of the classification task.

We select an $L_{1,2}$ regularization constraint to each level of the tree hierarchy, and feature ranking is performed to select the most relevant features among them. $L_{1,2}$ regularization constraint is an unbiased estimation that results in a sparser and more computationally efficient solution to the minimization problem compared to L_1 regularization[18].

Applying the sparse representation of feature selection to a tree hierarchy, feature selection can be performed based on whether the weight matrix of the features is zero or not.

Let the weight matrix $\boldsymbol{\omega}_i$ of each layer be divided into the sum of two components: $\boldsymbol{W}_i$ and $\boldsymbol{D}_i$. Then the three parameters are substituted into the feature selection model of the sparse representation to obtain the minimization loss function of the ith layer with respect to $\boldsymbol{W}_i$ and $\boldsymbol{D}_i$:

$$J(\boldsymbol{W}_i, \boldsymbol{D}_i) = \sum_{j=1}^{m_i} \frac{1}{n_{i,j}} \left\| \boldsymbol{X}_{i,j} \boldsymbol{\omega}_{i,j} - y_{i,j} \right\|^2 + \lambda_1 \left\| \boldsymbol{W}_i \right\|_{1,2} + \lambda_2 \left\| \boldsymbol{D}_i^{\mathrm{T}} \right\|_{1,2}$$

$$s.t \qquad \boldsymbol{\omega}_i = \boldsymbol{W}_i + \boldsymbol{D}_i \qquad 22\backslash* \text{ MERGEFORMAT ()}$$

where $\boldsymbol{X}_{i,j}$ is the sample matrix of the ith layer belonging to the jth class, $\boldsymbol{\omega}_{i,j}$ is the corresponding weight vector of $\boldsymbol{X}_{i,j}$, $n_{i,j}$ is the number of samples in $\boldsymbol{X}_{i,j}$, $y_{i,j}$ is the label of the class corresponding to $\boldsymbol{X}_{i,j}$, $\| \|_{1,2}$ is the $L_{1,2}$ paradigm, and λ_1 and λ_2 are non-negative parameters controlling the regularization. More specifically, $\boldsymbol{\omega}_{i,j}$, $\boldsymbol{W}_i$ and $\boldsymbol{D}_i$ are:

$$\boldsymbol{\omega}_i = \left[\omega_{i,1}, \omega_{i,2}, \cdots, \omega_{i,m_i} \right] \in R^{n \times m_i} \qquad 33\backslash* \text{ MERGEFORMAT ()}$$

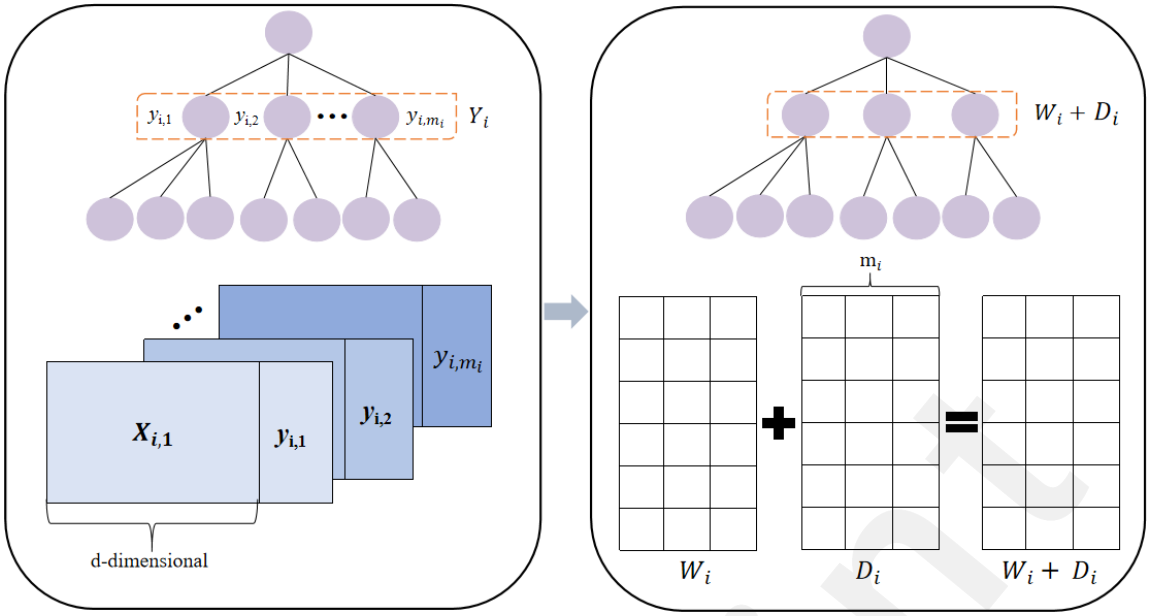
$$\boldsymbol{W}_i = \left[w_{i,1}, w_{i,2}, \cdots, w_{i,m_i} \right] \in R^{n \times m_i} \qquad 44\backslash* \text{ MERGEFORMAT ()}$$

$$\boldsymbol{D}_i = \left[d_{i,1}, d_{i,2}, \cdots, d_{i,m_i} \right] \in R^{n \times m_i} \qquad 55\backslash* \text{ MERGEFORMAT ()}$$

$$\omega_{i,j} = w_{i,j} + d_{i,j} \qquad 66\backslash* \text{ MERGEFORMAT ()}$$

where $i = 1, 2, \ldots, l$, l represents the number of levels in the tree hierarchy, $j = 1, 2, \ldots, m_i$, m_i represents the number of classes in the ith layer, and n represents the sample feature dimension.

Traditional hierarchical feature selection considers different nodes as independent and selects completely different subsets of features, whereas in this paper, we propose selecting a portion of the same feature for every layer, known as shared features identified by $\boldsymbol{W}_i$, for different nodes in the same layer. Additionally, we select exclusive features for each node that are suitable for classification identified by $\boldsymbol{D}_i$. Figure 2 illustrates the process of selecting shared features and different exclusive features based on the dataset at ith layer in the hierarchy.
Figure 2. Feature selection process at the ith layer



The procedure for solving the minimized loss function (1) is as follows: decompose equation (1) into $h(\boldsymbol{W}_i, \boldsymbol{D}_i)$ and $r(\boldsymbol{W}_i, \boldsymbol{D}_i)$:

$$h(\boldsymbol{W}_i, \boldsymbol{D}_i) = \sum_{j=1}^{m_i} \frac{1}{n_{i,j}} \left\| \boldsymbol{X}_{i,j} \left(\boldsymbol{W}_{i,j} + \boldsymbol{D}_{i,j} \right) - y_{i,j} \right\|^2,$$

$$r(\boldsymbol{W}_i, \boldsymbol{D}_i) = \lambda_1 \left\| \boldsymbol{W}_i \right\|_{1,2} + \lambda_2 \left\| \boldsymbol{D}_i^{\mathrm{T}} \right\|_{1,2} \quad 77\backslash * \text{ MERGEFORMAT ()}$$

where $h(\boldsymbol{W}_i, \boldsymbol{D}_i)$ is the empirical loss function and $r(\boldsymbol{W}_i, \boldsymbol{D}_i)$ is the regularization term. That is, the objective minimization function (1) can be expressed as the differential term $h(\boldsymbol{W}_i, \boldsymbol{D}_i)$ with non-differential $r(\boldsymbol{W}_i, \boldsymbol{D}_i)$.

Then Eq. (6) is expressed using Taylor's formula[20] and solved iteratively with the accelerated gradient descent[21], which converges after k iterations to obtain the optimal solution $\boldsymbol{W}_i^{(*)}, \boldsymbol{D}_i^{(*)}$. Ultimately, according to $\boldsymbol{W}_i^{(*)}, \boldsymbol{D}_i^{(*)}$, we obtain the weight matrix $\boldsymbol{\omega}_i^{(*)}$. The top N features are selected based on the ordering of the values of the elements in $\boldsymbol{\omega}_i^{(*)}$.

Based on the above description and definitions, Algorithm 1 describes the process of hierarchical feature selection at each level based on $L_{1,2}$ paradigm regularization.

Algorithm 1 Hierarchical feature selection based on $L_{1,2}$ paradigm regularization

Input: sample matrix $\boldsymbol{X}$, hierarchical tree of classes
Output: the top N features of each layer in the hierarchy
for i=1 to l-1 do
  $t_0 = 1, k = 1$
  while k<max do
    $t_k = (1 + \sqrt{t_{k-1}^2 + 1} / 2)$
    update $W_i^{(*)}$ and $D_i^{(*)}$ by accelerated gradient descent
    $k = k + 1$
  end
  produce the optimal solution $W_i^{(*)}$ and $D_i^{(*)}$ after k iterations, and then get $\omega_i^{(*)}$
  sort the values of each element in $\omega_i^{(*)}$ to obtain a feature ordering for each class in this layer

end
return: Feature ordering for each layer

Multi-path hierarchical classification

From the framework, it can be seen that after the hierarchical feature selection based on $L_{1,2}$ paradigm regularization, these features are used as the feature subset for classification. Subsequently, multi-path hierarchical classification is performed within the hierarchy. Assume that the sample matrix of the dataset is represented by $X \in \mathrm{R}^{m \times d}$, where m is the number of samples in the dataset and d is the selected data dimensions. Additionally, let $Y_i = \{y_{i,1}, y_{i,2}, ..., y_{i.m_i}\}$ denote the set of classes in the ith granularity layer, where m_i represents the number of granularity classes in that layer. The k pre-selected multipaths in this paper are limited to $1 \le k \le \min(m_i)$.

In the multi-path hierarchical classification process, the logistic regression is selected for the classifier, for it having many advantages: it is simple to implement, parallelizable, computationally inexpensive, fast, and can directly model the likelihood of classification. In addition, it not only predicts the classes but also provides the predicted probabilities for different classes[22]. The probability for the jth class in a given ith classification subtask is:

$$p_{i,j} = \frac{1}{1 + e^{-\omega_j^{\mathrm{T}} x - b_j}} \qquad 88\backslash * \text{ MERGEFORMAT ()}$$

where the ω_j represents the weight of the feature in the sample data, and b_j represents the bias. The weight(ω_j) and the bias(b_j) of the classification sub-tasks in each layer can be obtained by training on the training set. Usually, in tree hierarchies, the root node of the first layer generates all the nodes of the second layer. Therefore, the probability of the class nodes in the second layer can be calculated directly from equation mentioned above.

First of all, considering the class inclusion relationship present in the upper and lower layers of the tree hierarchy, the probability of each subclass in each layer of the hierarchy is calculated sequentially. According to the logistic regression model and the weights and biases obtained from the training, the formula for the probability value of the hth subclass $c_{j,h}$ generated by the jth class on the (i-1)th layer at the ith layer is obtained as:

$$q(x)_{i,c_{j,h}} = \frac{1}{1 + e^{-\omega_{c_{j,h}}^{\mathrm{T}} x - b_{cj,h}}}, c_{j,h} \in C_j \qquad 99\backslash * \text{ MERGEFORMAT ()}$$

where C_j is the set of all subclasses of the jth class.

The probability values P_{i,C_j} for each class in the ith layer are updated based on the parent-child class dependency between the ith granularity layer and its previous the (i-1)th granularity layer, as depicted in Equation (9):

$$P_{i,C_j} = \left(p(x)_{i-1,j} \times q(x)_{i,c_{j,h}} \right), 1 \le h \le C_j \qquad 1010\backslash * \text{ MERGEFORMAT ()}$$

where $1 \le h \le C_j, c_{j,h} \in C_j$.

Secondly, P_{i,C_j} is utilized as the foundation for class ranking. In each layer, the top k pre-selected

classes $Y_{i,k}$ are chosen based on their order until the top k pre-selected classes in the leaf node layer are acquired. The specific process is as follows: assuming that k preselected classes are selected from the previous (i-1)th layer, the subclasses generated by these k preselected classes in the ith layer can be obtained based on the relationship between the upper and lower layers. Subsequently, all the k^2 classes that are candidates to become preselected classes at the current ith layer can be obtained by the following equation (10):

$$P_{i,k_{C_1}} \cup P_{i,k_{C_2}} \cup \cdots P_{i,k_{C_{k-1}}} \cup P_{i,k_{C_k}} \Rightarrow P_{i,k^2}$$
$$Y_{i,k_{C_1}} \cup Y_{i,k_{C_2}} \cup \cdots Y_{i,k_{C_{k-1}}} \cup Y_{i,k_{C_k}} \Rightarrow Y_{i,k^2}$$
$$P_{i,k_{C_i}} = \max_k (P_{i,C_i})$$
$$Y_{i,k_{C_i}} = \max_k (Y_{i,C_i})$$

1111* MERGEFORMAT ()

where $P_{i,k_{Cj}}$ represent the set of probabilities of the k preselected subclasses generated by the subclasses of class j in the (i-1)th layer at the ith layer. Similarly, P_{i,k^2} represent the probabilities of the k^2 preselected subclasses on the ith layer. $Y_{i,k_{C_j}}$ represent the set of the k preselected subclasses generated by the subclasses of class j in the (i-1)th layer at the ith layer, while Y_{i,k^2} denotes the k^2 preselected subclasses on the ith layer.

The classes in the current ith layer are ranked based on the probability of occurrence among the k^2 possible classes. The top k classes are then selected to form the pre-selected classes $Y_{i,k}$ for the ith granularity layer. This process can be described as:

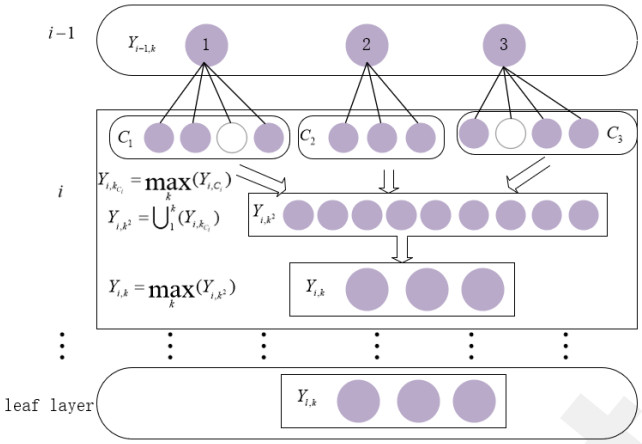
$$P_{i,k} = \max_k \left(P_{i,k^2} \right)$$

1212* MERGEFORMAT ()

Sequentially, starting from the top and moving downwards, the k pre-selected classes at the (i-1)th granularity layer result in k^2 possible classes at the ith granularity layer. These are followed by the top k pre-selected classes with the highest probability values. Lastly, the k pre-selected classes are chosen at the leaf node layer.

Figure 3 details how the multipath selection process is performed at each layer(k=3). The three pre-selected classes chosen from the (i-1)th layer generate a set of three subclasses, C_1, C_2, C_3, at the ith layer. By selecting the first three classes according to their probability values, a set of nine candidate pre-selected classes is obtained. The selection of the three pre-selected classes in each layer is determined by the probability values assigned to the nine classes. This approach enables a multi-path hierarchical classification process, which proceeds from the top to the bottom layers. Ultimately, the three pre-selected classes in the leaf layer are identified.

Figure 3. Process of pre-selection k=3 pre-selection classes in hierarchical classification



Finally, a classifier is employed to forecast the ultimate classification within the k pre-selected classes derived from the multipath results. The optimal outcome is achieved when the true class of the predicted samples falls within one of the k pre-selected classes, and the base classifiers accurately classify among the k minority classes. Algorithm 2 outlines the process of multi-path hierarchical classification.

Algorithm 2: Multi-path hierarchical classification

Input: sample feature subset
Output: predict class
for i=1 to l do
    train the feature weights and biases of each class for each subclassification task at each layer using logistic regression in the test set
end
for i=1 to l do
    from the probability values of the preceding layer, the individual class probability values of the ith layer can be obtained using Eq. (8)
    for t=1:*k*
        get P_{i,k^2} and Y_{i,k^2} by Eq. (10)
        select $P_{i,k}$ and $Y_{i,k}$ of the ith layer by Eq. (11)
    end
end
determination of final prediction classes among pre-selected classes using base classifiers
return: predict class

Dataset

Info

The dataset utilized in our research is obtained from the clinical diagnostic records of patients with FUO who were admitted to the First Affiliated Hospital of Xi'an Jiaotong University between the years 2011 and 2020. Each sample in this study represents authentic clinical data obtained from patients with FUO, encompassing pathological data and diagnoses provided by physicians. The pathological data encompasses a range of information, including clinical symptoms, epidemiological history, past medical history, laboratory tests, medical imaging, and indicators from pathological examination. The statistical indicators of the dataset are presented in Table 1.

Table 1. Statistical Analysis Results of Some Indicators

indicators	number of samples
sex	M/F=303/261
age	0~20/20~40/40~60/60~=87/172/188/117
infectious	399
noninfectious	165

For the purpose of this study, we utilized patients' pathological data and doctors' diagnostic results as the training dataset in order to construct the model. Due to the limited quantity of available data, there is a possibility of encountering a significant imbalance within the dataset. This imbalance may result in a bias towards predicting classes that have larger data samples, ultimately impacting the overall classification performance. During the process of data analysis, samples that contained less than six instances of a particular disease were excluded in order to address the issue of imbalance. After undergoing the refinement process, a final dataset consisting of 564 samples was obtained. This dataset encompasses 16 exact classes of etiology for FUO. Please consult Table 2 for additional information regarding the dataset.

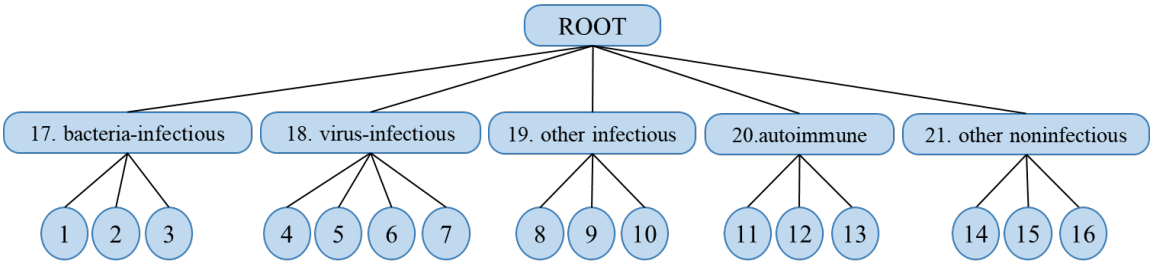
Table 2. Dataset Summary Results

	Diagnose	Number	Label
Bacteria-infectious			
	Liver abscess	24	1
	Endocarditis	12	2
	Brucellosis	64	3
Virus -infectious			
	Epstein-barr virus infection	77	4
	Cytomegalovirus infection	14	5
	Infectious mononucleosis	38	6
	Viral infection	103	7
Other infectious			
	Kala-azar	11	8
	Mycoplasma infection	11	9
	Rickettsia infection	45	10
Autoimmune			
	Anca-associated vasculitis	12	11
	Adult Onset Still’s Disease	20	12
	Lymphoma	33	13
Other noninfectious			
	Systemic inflammatory response syndrome	47	14
	Hemophagocytic syndrome	19	15
	Necrotizing lymphadenitis	34	16

Hierarchy label

According to the pathological characteristics of FUO, the dataset can be organized in a hierarchical tree structure[23], as depicted in Figure 4. The categories of the FUO tree span from abstract etiology to concrete etiology, progressing from the root node to the leaf nodes in a top-to-bottom manner. In the dataset, the hierarchical tree structure exhibits a three-tiered system of granularity. The final level consists of 16 etiology-specific categories, while the middle level is labelled with categories 17 to 21, representing the coarse classifications.

Figure 4. Hierarchical structure



Data preprocessing

The clinical symptoms, epidemiological history, past medical history, and laboratory tests in the pathological data of the FUO patients exhibited a range of different forms, including both continuous and discrete data. Therefore, the data underwent preprocessing to ensure standardization.

Imputation of missing values: In order to address missing values in the dataset, the KNN method was employed for interpolating continuous missing data[24]. For discrete missing data, the approach involved utilizing the mode of all available data points within the respective data item for filling the missing values.

Coding: After filling in the missing values in the discrete data, it was necessary to identify the category features that lack significance in terms of size. Subsequently, numerical or vectorization operations can be applied to these features. Two coding methods, namely 0-1 vector and one-hot vector, were employed in the paper.

Normalization: A process applied to continuity data after filling in missing values, aiming to distribute the values on the [0,1] interval, to result in the pre-processed normalized data[25]:

$$x' = \frac{x - \mathrm{MIN}}{\mathrm{MAX} - \mathrm{MIN}} \qquad 1313\backslash* \text{ MERGEFORMAT ()}$$

where x represents the original continuity data, MIN represents the minimum value of the data item within its respective location, and MAX represents the maximum value of the data item within its respective location.

After undergoing data preprocessing, the final dataset consists of 564 samples totally, with each sample having 327 dimensions.

Compared methods

To evaluate the performance of the proposed MPFSHC, experiments were conducted and compared with four similar hierarchical classification methods:

(1) TDLR: Top-Down Logistic Regression hierarchical classification. At each granularity level in the hierarchy, it selects the node with the highest predicted probability as the classification result, recursively performing from top to bottom until reaching the leaf level.

(2) HNBP[26]: Hierarchical classification based on optimal N-paths. The proposed approach converts the task of class prediction into a search problem, aiming to identify multiple paths within a tree-like hierarchy that have the highest joint probability. This strategy effectively mitigates the issue of error propagation between different levels.

(3) CSHCIC[27]: Cost-sensitive hierarchical classification based on class hierarchy correlation. In the same layer of hierarchical classification, there is an imbalanced data distribution,

introducing cost-sensitive factors to reduce the tendency of majority class classification, improving the classification accuracy of minority classes.

(4) CSHC[28]: Cost-sensitive hierarchical classification based on multi-scale information entropy. The computation of information entropy for various classes at each level of the hierarchy is performed, and an entropy threshold is established to mitigate the propagation of errors from higher-level classification tasks to lower-level ones. It assigns different cost weights to classes based on hierarchical information entropy to address data imbalance.

Evaluation metrics

The performance of the proposed method was assessed and confirmed through a series of experiments. Five metrics were used for evaluation: Hierarchical F_1-measure(F_{H})[29], F_{H} based on the losest ancestor(F_{LCA})[30], Tree Induced Error(TIE)[31], Accuracy(ACC) and Runtime (T).

(1) Hierarchical F_1-measure(F_{H}): Let Y denote the true label of sample X, $\hat{Y}$ represent the predicted label of X, and $anc(Y)$ denote the set of parent nodes that accurately predicted samples. The hierarchical accuracy P_H, hierarchical recall R_H and F_H applicable to hierarchical structures are defined as follows:

$$P_H = \frac{\left|\hat{Y}_{aug} \cap Y_{aug}\right|}{\left|\hat{Y}_{aug}\right|}$$

$$R_H = \frac{\left|\hat{Y}_{aug} \cap Y_{aug}\right|}{\left|Y_{aug}\right|}$$

$$F_H = \frac{2 \cdot P_H \cdot R_H}{P_H + R_H} \qquad 1414\backslash * \mathrm{MERGEFORMAT}\ ()$$

where $Y_{aug} = Y \cup \mathrm{anc}(Y)$, $\hat{Y}_{aug} = \hat{Y} \cup \mathrm{anc}(\hat{Y})$, $|\square|$ represented as the number of sets.

(2) F_{H} based on the losest ancestor(F_{LCA}): The Lowest Common Ancestor (LCA) in the context of a tree-based hierarchical structure refers to the node that is the deepest and furthest from the root node among the common ancestors of the true class node and the predicted class node. Let C_{aug} represent the set of nodes that lie along the path from the true class to the LCA node in the tree structure, and let $\hat{C}_{aug}$ represent the set of nodes along the path from the predicted class to the LCA node. The accuracy P_H (P_{LCA}) and recall R_H (R_{LCA}) metrics based on LCA are defined as follows:

$$P_{LCA} = \frac{\left|\hat{C}_{aug} \cap C_{aug}\right|}{\left|\hat{C}_{aug}\right|}$$

$$R_{LCA} = \frac{\left|\hat{C}_{aug} \cap C_{aug}\right|}{\left|C_{aug}\right|}$$

$$F_{LCA} = \frac{2 \cdot P_{LCA} \cdot R_{LCA}}{P_{LCA} + R_{LCA}} \qquad 1515\backslash * \mathrm{MERGEFORMAT}\ ()$$

F_H metric counts multiple common ancestors, including the root node, which makes it difficult to accurately distinguish between different degrees of error at the lower level. On the other hand, F_{LCA} metric only considers subtrees rooted in the actual node class and the LCA of the predicted node. This allows for a more precise comparison of the differences in splitting errors among the lower

nodes. The larger F_{LCA} value indicates that the fewer paths passed from the real class nodes to the predicted class nodes, the lower the degree of errors, and the better the classification results.

(3) Tree Induced Error(TIE): In the process of classifying a hierarchical structure, varying degrees of prediction errors lead to distinct penalties. The TIE is a measure of the distance between the predicted class and the true class of a tree, and is defined as follows:

$$\mathrm{TIE}(Y,\hat{Y}) = \left|E_{\mathrm{h}}(Y,\hat{Y})\right| \qquad 1616\backslash * \mathrm{MERGEFORMAT}\ ()$$

where $E_{\mathrm{h}}(Y,\hat{Y})$ represents the set of edges that are traversed from the true class node Y to the predicted class node $\hat{Y}$ within a tree-like hierarchical structure, $|\square|$ denotes the cardinality of this set, which refers to the number of edges.

(4) Accuracy(ACC):

$$\mathrm{ACC} = \frac{R_N}{N} \qquad 1717\backslash * \mathrm{MERGEFORMAT}\ ()$$

where R_N represents the count of correct predictions, while N represents the total number of samples in the test set.

(5) Runtime (T)
Runtime encompasses both the duration of the training phase and the testing phase.

Results

Ablation experiments

The validation of the number of hierarchical paths

Our study conducted a comparison of the performance of pre-selected classes using various values of k. Within the context of the hierarchical classification method proposed in this paper, in order to assess the performance of different k-values while maintaining consistency in other parameters, a feature selection approach was employed, selecting 100% of the features. We chose SVM and random forest [32] as the base classifiers and named the two schemes MPFSHC-SVM and MPFSHC-RF, respectively. Empirical findings are presented in Table 3.

Table 3. Performance comparison of different k values in the pre-selected class

	k	ACC(%)	F_{H}(%)	F_{LCA}(%)	TIE	T(s)
MPFSHC-SVM						
	1	66.66	82.03	79.90	60.8	**0.87**
	2	71.49	84.49	82.74	52.6	2.18
	3	**72.35**	**85.01**	**83.29**	**50.8**	4.17
	4	71.83	84.48	82.85	52.6	7.84
	5	68.47	82.77	80.87	58.4	10.00
MPFSHC-RF						
	1	66.66	82.03	79.90	60.8	**3.31**
	2	68.97	83.05	81.18	57.4	43.20
	3	**69.20**	**83.25**	**81.36**	**56.8**	45.19
	4	64.90	80.45	78.53	66.2	50.63

	5	65.08	80.15	78.44	67.2	50.94

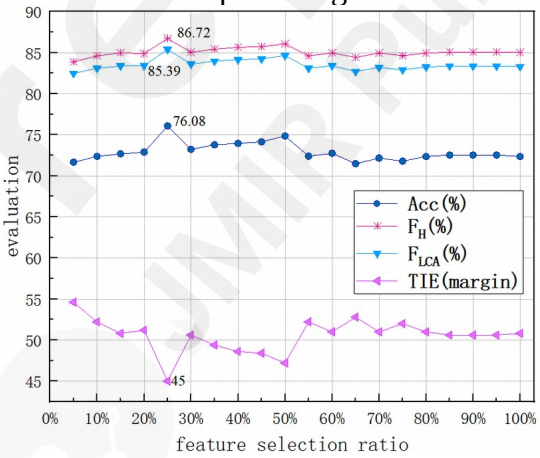
It is evident that varying k-values yield distinct performance metrics. Both MPFSHC-SVM and MPFSHC-RF achieve the best overall performance when k=3, which indicates that retaining multiple pre-selected classes can weaken the error propagation problem of the hierarchical classification to a certain extent, but it also indicates that the larger the value of k is, the not necessarily better it is. In terms of time metrics, T, the runtime for k=3 exhibits an increase compared to that of k=1. This can be attributed to the fact that the number of paths increases, leading to a longer decision time for the system. However, it is important to note that the runtime for k=3 still falls within an acceptable range.

Compared to utilizing RF as the underlying classifier, MPFSHC-SVM demonstrates superior performance across all performance metrics. This is attributed to the superior performance of SVM in handling small sample sizes, nonlinearity, and high-dimensional classification problems. These characteristics align with the specific attributes of the FUO dataset that were investigated in our study.

The validation of different percentages of feature selection

A comparison was conducted to evaluate the performance of various feature selection percentages. The selection of features at each level of the hierarchical tree structure was consistent, with an equal percentage being chosen. From the Figure 5, it is evident that there is a noticeable trend as the feature selection percentage increases from 5% to higher values. Specifically, ACC, F_{H} and F_{LCA} all exhibit an upward trend, while TIE shows a decreasing trend. The curves for ACC, F_{H} and F_{LCA} exhibit similar trends, while the trend of TIE is opposite to these curves. When the feature selection reaches 25%, the highest level of performance is attained, with ACC of 76.08%, F_{H} of 86.72%, F_{LCA} of 85.39%, and TIE reduced to 45.

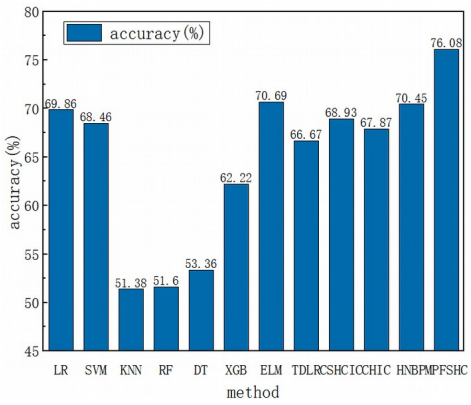
Figure 5. Performance comparison of different percentage feature selection



Comparison experiments

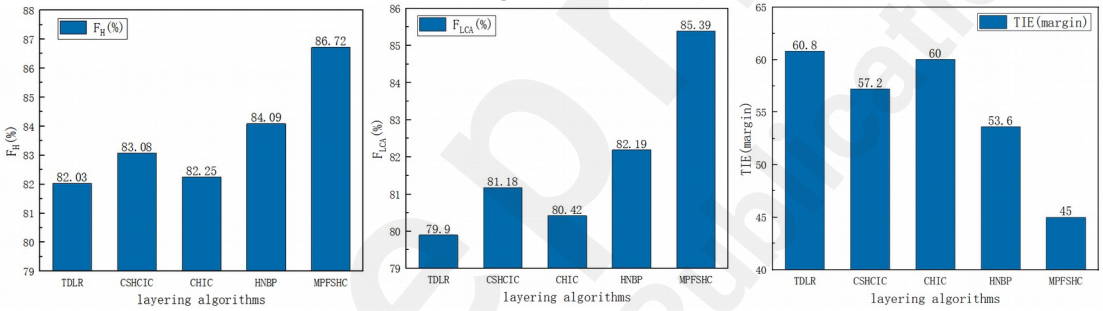
A comparison was conducted to assess the performance differences between the proposed method and alternative approaches. The proposed method utilized a value of 3 for the number of paths (k), employed SVM as the base classifier, and set the feature selection percentage to 25%.

Figure 6. ACC of different classification algorithms



In order to evaluate the accuracy of the assessment, a variety of comparison methods were employed, including both hierarchical classification techniques and flat classification techniques such as logistic regression (LR), KNN, random forest (RF)[32], SVM, extreme gradient boosting (XGB)[33], and extreme learning machine (ELM)[34], as depicted in Figure 6. The results clearly indicate that LR, SVM and ELM demonstrated relatively high performance compared to other flat classification methods, while the remaining flat methods exhibited lower accuracy. On the contrary, the hierarchical classification methods demonstrated strong performance, with the proposed method exhibiting the highest level of effectiveness, surpassing all other alternative approaches.

Figure 7. Validation of different hierarchical algorithms(a)F_H;(b) F_{LCA};(c)TIE

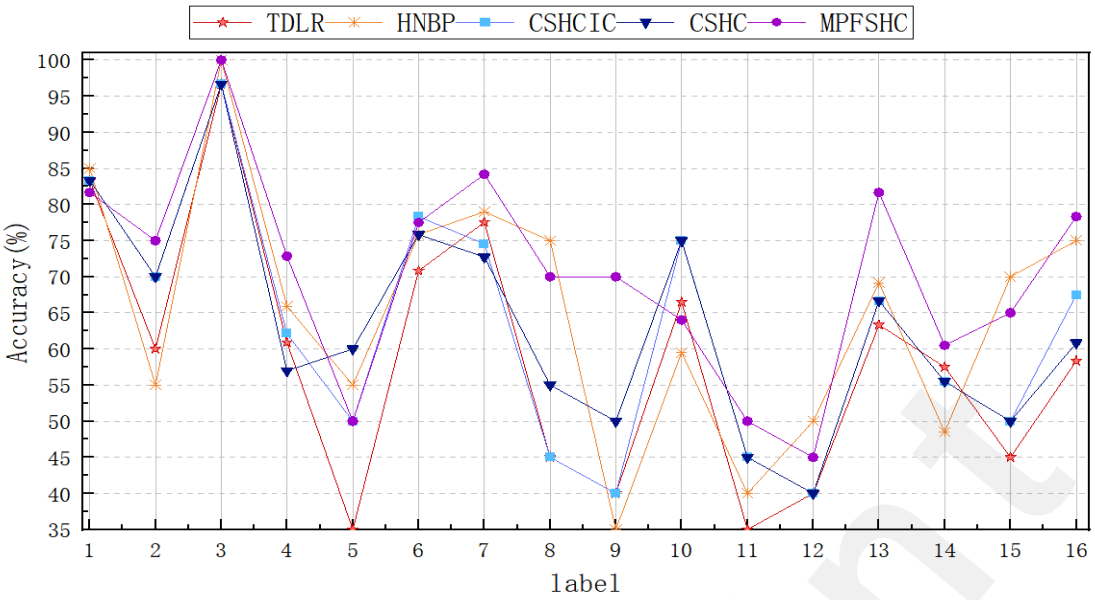


(a) (b) (c)

In relation to the F_H and F_{LCA} metrics, the proposed method was compared to other hierarchical classification methods, and the outcomes are depicted in Figure 7(a)and(b). The results indicate that MPFSHC achieved the highest rankings in both metrics. It obtained F_H of 86.72%, which is 2.63% higher than the second-ranked HNBP, and achieved F_{LCA} of 85.39%, surpassing HNBP by 3.2%. Figure 7(c) illustrates the TIE attained by various hierarchical classification methods. The results indicate that our method demonstrated the lowest TIE with a value of 45, representing a significant decrease of 8.6 compared to the second-ranked HCMP. This observation serves as evidence that our approach exhibits fewer misclassifications and superior classification performance.

To evaluate the classification performance of various algorithms on different classes, Figure 8 presents a comparison of the classification outcomes of MPFSHC with other hierarchical classification algorithms at the leaf layer for each individual class. Based on the obtained results, it is evident that the MPFSHC algorithm demonstrates superior classification accuracy across the majority of classes.

Figure 8. The classification accuracy of each method in each class of the FUO leaf layer



Additionally, we compared the accuracy of the coarse classes at the intermediate level of the FUO dataset, as shown in Table 4. It is evident that the proposed algorithm MPFSHC outperforms other hierarchical classification algorithms in terms of accuracy across all five coarse classes. The aforementioned observation suggests that the implementation of MPFSHC successfully addresses the issue of error propagation. In the context of class 17, the MPFSHC demonstrates a prediction accuracy of approximately 98%. In class 20, the MPFSHC demonstrates the most significant improvement in comparison to other methods, exhibiting a 15% improvement over the TDLR and an almost 8% improvement over the HNBP. Despite the relatively low overall prediction accuracy observed in all methods for classes 19-21, the MPFSHC algorithm demonstrates a notable improvement in the prediction accuracy of these three classes, surpassing 70%. This finding suggests that the proposed algorithm significantly enhances the predictive performance. From the analysis of the accuracy metrics, it is apparent that there exist misclassified test samples. This can be attributed to several factors, including the imbalanced distribution of samples across different classes in the dataset, the inherent variability present in the sample data, and the inadequate cleaning of the sample data resulting in sample overlap.

Table 4 Classification accuracy of each method on the coarse class of the middle layer in FUO(%)

label	TDLR	HNBP	CSHCIC	CSHC	MPFSHC
17	96.73	94.55	94.61	93.82	**97.96**
18	87.05	88.95	89.50	89.38	**89.72**
19	66.67	65.03	67.06	65.01	**71.18**
20	61.57	68.86	62.90	60.17	**76.75**
21	68.28	71.68	68.53	68.52	**74.75**
Mean	76.06	77.81	76.52	75.38	**82.07**

Discussion

Principal Results

From the results of our method, the case of k=1 in Table 3 is equivalent to using the traditional single-path hierarchical classification method. In this case, the ACC, F_H, F_{LCA}, and TIE metrics using SVM are 66.66%, 82.03%, 79.90%, and 60.8, respectively. The performance is the lowest among the results for different numbers of paths, as evidenced by the highest TIE. However, the T of 0.87 seconds is the shortest for this case, thanks to the single-path hierarchical approach that simplifies the model. When k=5, this scenario is equivalent to directly flattening the dataset for classification, as the dataset in this paper only has five coarse categories. The ACC of the MPFSHC -SVM is 68.47%,

which aligns closely with the SVM outcomes of various classification algorithms shown in Fig. 6, thus validating the earlier inference. Although in this case, the ACC of MPFSHC -RF is 13.48% higher than that of Fig. 6, it is due to the random nature of the classification mechanism of RF. In contrast, the optimal hyperplane sought by SVM is constrained by the spatial distribution of the samples. Consequently, the outcomes of each search are not significantly varied. Therefore, this discrepancy does not impact the conclusion that it is comparable to the direct flat classification of the dataset in the previous instance k=5. By comparing the classification results of multiple paths, it can be observed that both MPFSHC -SVM and MPFSHC -RF exhibit the best performance when k=3. The ACC of MPFSHC -SVM is 72.35%, representing an improvement of 5.69% and 3.88% compared to the single-path hierarchical classification with k=1 and the similar flattened classification with k=5, respectively. The ACC of MPFSHC -RF is 69.20%, showing an improvement of 2.54% and 4.12% over the two approaches mentioned above. The results of both classifiers demonstrate that the multi-path hierarchical classification approach can reduce the inter-layer error propagation problem. Additionally, decomposing the total task into multiple sub-tasks can reduce the complexity of the problem and improve the classification results. The running times of SVM and RF are 4.17 s and 45.19 s, respectively. These times are 3.3 s and 41.88 s longer than that of the single-path hierarchical classification, respectively. This suggests that more paths will increase the complexity of the hierarchical model, prolonging the decision time of the system. However, it is still within an acceptable range.

After incorporating feature selection based on hierarchical classification, it can be observed from Fig. 5 that the optimal classification performance is achieved when the proportion of feature selection is 25%. The ACC, F_{H}, and F_{LCA} metrics are 76.08%, 86.72%, and 85.39%, which are 3.73%, 1.71%, and 2.10% higher than those before the feature selection. Additionally, the TIE is 45, reduced by 5.8, indicating that with the appropriate proportion of screening features or number, the model's performance can be enhanced.

Compared with other methods, our method achieved the most optimal results, as demonstrated in Fig. 6 and Fig. 7. In the 16 classes of etiology our method is the most optimal among the four hierarchical classifications, outperforming other algorithms in nine classes. Specifically, they are classes of 2, 3, 4, 7, 9, 11, 13, 14, and 16. In the specific classes of 7, 9, and 13, our method demonstrates a significant superiority over the other four comparison algorithms, improving accuracy by up to 10%. Moreover, it achieved an accuracy of 100% in the class of 3. Regarding the five coarse categories (Table 4), our method's accuracy is 82.07%, showing improvements of 6.01%, 4.26%, 5.55%, and 6.69% over TDLR, HNBP, CSHCIC, and CSHC, respectively. This improvement is notably significant.

Conclusions

This paper presents a diagnostic method for Fever of Unknown Origin (FUO) utilizing multi-path feature selection and hierarchical classification (MPFSHC). Firstly, a hierarchical structure is constructed to identify the causes of FUO. A classification method is proposed in order to address the issue of inter-level error propagation in hierarchical classification, involving the pre-selection of multiple paths based on hierarchical prediction. Additionally, the $L_{1,2}$ regularization constraint is employed at each level within the hierarchical structure to facilitate feature selection. The objective is to eliminate redundant and interfering features, thereby enhancing the overall performance of the method. Experimental findings indicate that the implementation of a hierarchical classification model significantly enhances the accuracy of predicting FUO. Moreover, the incorporation of multiple path selection and feature selection further amplifies the effectiveness of the hierarchical classification model, offering a potential

direction for the intelligent diagnosis of FUO.

Acknowledgements

Jianchao DU and Junyao DING contributed to the method design. Jianchao DU, Junyao DING and Yuan WU contributed to in review and approval of the manuscript. Tianyan CHEN and Lei SHI contributed to the acquisition of the data set. Jianqi LIAN and Yun ZHOU are role of sponsors.
This research was supported by the Second Affiliated Hospital of Air Force Medical University (2021QYJC-005).

Conflicts of Interest

None declared.

Abbreviations

FUO: fever of unknown origin
ML: machine learning
MPFSHC: hierarchical classification method based on multi-path and feature selection
TDLR: top-down logistic regression hierarchical classification
HNBP: hierarchical classification based on optimal N-paths
CSHCIC: cost-sensitive hierarchical classification based on class hierarchy correlation
CSHC: cost-sensitive hierarchical classification based on multi-scale information entropy
SVM: support vector machine
RF: Random Forest
LR: Logistic Regression,
KNN: K-Nearest Neighbor
XGB: Extreme Gradient Boosting,
ELM: Extreme Learning Machine

References

1. Petersdorf RG, Beeson PB. Fever of unexplained origin: report on 100 cases. Medicine 1961 Feb;40(1):1-30. [Medline]
2. Mulders-Manders C, Simon A, Bleeker-Rovers C. Fever of unknown origin. Clin Med (Lond) 2015 Jun;15(3):280-284. [Medline]
3. Romy Younan, Laure Yammine, Claude Afif, et al. Isolated Bone Marrow Infiltration by Streptococcus Pneumoniae: An Unusual Etiology of Fever of Unknown Origin. Clinical Lymphoma Myeloma and Leukemia 2022 Oct; 22(S2): S441-S442. [CrossRef]
4. Kosuke Ishizuka, Takanori Uehara, Makoto Arai, et al. Medical-type peritoneal mesothelioma leading to death two months after onset of fever of unknown origin. Radiology Case Reports 2022 Dec; 17(3):540-543. [CrossRef]
5. Mani Bhushan Kumar, Lokesh Varada, Kishore Abuji, et al. Mycotic aneurysm by Bacteroides fragilis presenting as fever of unknown origin. Indian Journal of Medical Microbiology 2023 Jan-Feb; 41:53-54. [CrossRef]
6. Yan Yongjie, Chen Chongyuan, Liu Yunyu, et al. Application of Machine Learning for the Prediction of Etiological Types of Classic Fever of Unknown Origin. Frontiers in Public Health 2021 Dec; 9: 1-11. [Medline]
7. Okuducu YK, Nwosu A, Awad A, Ratna BB. Fever of Unknown Origin in a 17-Year-Old Girl. Cureus 2020 Sep; 12(9): e10607. [Medline]
8. Peng F, Wang H, Zhuang L, et al. Methods of enterprise electronic file content information mining under big data environment. 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE);2020 Oct 30-Nov 1; Bangkok, Thailand.

IEEE;2021. [CrossRef]
9. Kaur P, Sharma A, Chahal JK, et al. Analysis on Credit Card Fraud Detection and Prevention using Data Mining and Machine Learning Techniques. 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA);2021 Nov 26-27; Nagpur, India. IEEE;2022. [CrossRef]
10. Sang Q, Dai J, Tu S. Coal Mine Safety Risk Prediction Based on Incremental Extreme Learning Machine. 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC); 2022 Apr 14-16; Dalian, China. IEEE;2022. [CrossRef]
11. CHOUDHURY A. Identification of Cancer--Mesothelioma Disease Using Logistic Regression and Association Rule. arXiv Preprint posted online August 21, 2019. [FREE Full text]
12. AL-TASHI Q, RAIS H, JADID S. Feature selection method based on grey wolf optimization for coronary artery disease classification. International conference of reliable information and communication technology. Cham: Springer:2018. [CrossRef]
13. PAWLOVSKY AP. An ensemble based on distances for a kNN method for heart disease diagnosis. 2018 international conference on electronics, information, and communication (ICEIC);2018 Jan 24-27; Honolulu, HI, USA. IEEE; 2018. [CrossRef]
14. OGUNLEYE A, Wang QG. XGBoost model for chronic kidney disease diagnosis. IEEE/ACM transactions on computational biology and bioinformatics 2020 Nov-Dec; 17(6): 2131-2140. [Medline]
15. POLAT K. A hybrid approach to Parkinson disease classification using speech signal: the combination of smote and random forests.2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT); 2019 Apr 24-26; Istanbul, Turkey. IEEE;2019. [CrossRef]
16. Yelure BS, Patil SV, Patil SB, et al. Solving Multi-Class Classification Problem Using Support Vector Machine. 2022 International Conference on Futuristic Technologies (INCOFT); 2022 Nov 25-27; Belgaum, India. IEEE; 2023. [CrossRef]
17. Deng L, Sui Y, Chen L, et al. Hierarchical Classification Boost Using Confidence Belief Propagation. 2020 IEEE 6th International Conference on Computer and Communications (ICCC); 2020 Dec 11-14; Chengdu, China. IEEE: 2021. [CrossRef]
18. XU Zongben, GUO Hailiang, WANG Yao, et al. Representative of L1/2 regularization among Lq ($0 < q \leq 1$) regularizations: an experimental study based on phase diagram. Acta Autom Sin 2012 July; 38(7): 1225-1228. [CrossRef]
19. FREEMAN C, KULIC D, BASIR O. Feature-selected tree-based classification. IEEE transactions on cybernetics 2013 Dec; 43(6): 1990-2004. [CrossRef]
20. Wang X, Li Y, Wang D. The PSO and the Taylor Expansion Based LSSVM Algorithm for EEG Fatigue Multi-Classification. 2023 IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT); 2023 Jul 21-24; Qingdao, China. IEEE: 2023. [CrossRef]
21. Galetto FJ, Deng G. Reverse image filtering using total derivative approximation and accelerated gradient descent. IEEE Access 2022 Nov; 10: 124928-124944. [CrossRef]
22. Peng CYJ, Lee KL, Ingersoll GM. An introduction to logistic regression analysis and reporting. The journal of educational research 2002 Sep; 96(1): 3-14. [CrossRef]
23. Silla, C.N., Freitas, A.A. A survey of hierarchical classification across different application domains. Data Min Knowl Disc 2011 Jan; 22:31-72. [CrossRef]
24. Beretta L, Santaniello A. Nearest neighbor imputation algorithms: a critical evaluation. BMC medical informatics and decision making 2016 Jul; 16(3): 74. [Medline]
25. Izonin I, Ilchyshyn B, Tkachenko R, et al. Towards data normalization task for the efficient mining of medical data. 2022 12th International Conference on Advanced Computer Information Technologies (ACIT); 2022 Sep 26-28; Ruzomberok, Slovakia. IEEE; 2022.

[CrossRef]
26. QU Yanyun, LIN Li, SHEN Fumin, et al. Joint hierarchical class structure learning and large-scale image classification. IEEE Transactions on Image Processing 2016 Oct; 26(9): 4331-4346. [CrossRef]
27. ZHENG Weijie, ZHAO Hong. Cost-sensitive hierarchical classification for imbalance classes. Applied Intelligence 2020 Mar; 50(8): 2328-2338. [CrossRef]
28. ZHENG Weijie, ZHAO Hong. Cost-sensitive hierarchical classification via multi-scale information entropy for data with an imbalanced distribution. Applied Intelligence 2021 Jan; 51: 5940-5952. [CrossRef]
29. GOMEZ J C, MOENS M F. Hierarchical classification of web documents by stratified discriminant analysis. Multidisciplinary Information Retrieval: 5th Information Retrieval Facility Conference; 2012; Vienna, Austria. Berlin Heidelberg: Springer. [CrossRef]
30. SCHIEBER B, VISHKIN U. On finding lowest common ancestors: Simplification and parallelization. Lecture Notes in Computer Science; 1988; New York, USA. New York: Springer; 2005. [CrossRef]
31. DEKEL O, KESHET J, SINGER Y. Large margin hierarchical classification. Proceedings of the twenty-first international conference on Machine learning; 2004 Jul 4-8; Alberta, Canada. New York: Association for Computing Machinery; 2004. [CrossRef]
32. Leo Breiman. Random Forests. Machine Learning 2001 Oct; 45(1):5-32. [CrossRef]
33. Tianqi Chen, Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining; 2016 Aug 13-17; California, USA. New York: Association for Computing Machinery;2016. [CrossRef]
34. GuangBin Huang, QinYu Zhu, CheeKheong Siew. Extreme learning machine: Theory and applications. Neurocomputing 2006 Dec; 70(1–3):489-501. [CrossRef]