

DESCRIPTION

TITLE OF INVENTION: [SYSTEM FOR PREDICTING EPILEPSY BASED ON SYNCHRONOUS ACQUISITION OF EEG AND EGEG SIGNALS]

[0001] This application claims the priority of Chinese invention patent application CN2024113110412, titled "System for predicting epilepsy based on synchronous acquisition of EEG and EGEG signals", filed on September 19, 2024, the entire content of which is incorporated herein by reference.

TECHNICAL FIELD

[0002] The present invention relates to the field of disease prediction, and in particular to a system for predicting epilepsy based on synchronous acquisition of electroencephalogram (EEG) and electrogastroenterogram (EGEG) signals.

BACKGROUND

[0003] Epilepsy is a chronic neurological disease that affects approximately 70 million people worldwide. It has been listed by the World Health Organization as one of the five key neuropsychiatric disorders prioritized for global prevention and treatment, and is also a key research area of the Global Brain Initiatives. Epileptic seizures are caused by abnormal discharge of a group of brain cells, and different parts of the brain may become sites of the abnormal discharge. They are characterized by recurrence of seizures, abrupt seizure and abrupt cessation, and difficulty in capturing the EEG during the ictal period. During the epileptic seizures, transient involuntary convulsions often occur in a certain part of the body or the entire body, sometimes accompanied by loss of consciousness, as well as urinary and fecal incontinence. The seizures range from extremely transient loss of consciousness or muscle reflexes to severe and continuous convulsions. The frequency of the seizures can also vary, ranging from less than once a year to several times a day.

[0004] At present, the diagnostic basis for epilepsy includes clinical seizures and epileptiform discharges on an electroencephalogram. In an aspect, the evaluation of EEG abnormalities may be subjective and not very sensitive. For example, in routine

EEG recordings, only 29%-55% of patients have epileptiform discharges (such as spike waves, sharp waves, spike-and-slow wave complex, and sharp-and-slow wave complex). In another aspect, a large proportion of patients with epilepsy have seizures that are short in duration, making it difficult to capture abnormal discharges thereof in spite of performing EEG examination thereon. Therefore, the prior art based on the EEG can only detect epileptiform discharges in a subject at present, and is further suitable for the diagnosis of epileptic seizures, but difficult to serve to screen epilepsy.

[0005] According to reports, up to 70% of patients with epilepsy can achieve seizure freedom through proper use of antiepileptic drugs. As can be seen, “early detection and treatment” of epilepsy is of great practical significance. However, there are at least 10 million patients with epilepsy in China, but only about one-third of them are treated in a proper and adequate manner. The Global Burden of Disease Study estimates that the epilepsy burden in China is 1.6 million disability-adjusted life years (DALY), accounting for 12% in the world and 95% in the East Asian region. The reason for the above results may be the long-term misunderstanding, fear, and bias of patients with epilepsy towards epilepsy, which leaves a considerable number of people in China unwilling to go to medical institutions for further examination even if they have epileptic seizures; moreover, the existing diagnosis of epilepsy is based on the detection of epileptiform discharges in the subject, which causes patients with epilepsy who do not have the epileptic discharge during the EEG detection to be omitted by the prior art. For some other patients who may have non-epileptic paroxysmal disorders (such as psychogenic non-epileptic seizures and transient ischemic attacks), they often worry about having epilepsy, but it is difficult to distinguish from epilepsy through clinical manifestations alone. A correct diagnosis is the prerequisite for effective treatment, but in the prior art, it is difficult to distinguish between patients without non-epileptic seizures and patients with epilepsy just by virtue of patients’ oral statements and the EEG where abnormal discharges cannot be observed, especially in relatively large population such as communities and physical examination centers.

[0006] The Chinese patent application CN117481666A discloses a method for coupling an EGG signal with an EEG signal in a resting state, which reflects the degree of bidirectional influence between a raw EGG signal and a raw EEG signal by

extracting indicators of phase-amplitude coupling, phase-phase coupling, and amplitude-amplitude coupling between the raw EGG signal and the raw EEG signal in different frequency bands, as well as transfer entropy. However, the application scenarios thereof are limited to regulation of the gastric function, and it is difficult to diagnose diseases.

SUMMARY

[0007] The present invention provides a system for predicting epilepsy based on synchronous acquisition of EEG and EGEG signals, comprising:

[0008] a data acquisition module configured to acquire data comprising EEG data and EGEG data, wherein the EEG data is acquired from a first channel, the EGEG data is acquired from a second channel, and the data comprises sample data from sample population and subject data from a subject;

[0009] a data preprocessing module configured to preprocess the data to obtain corresponding time series data;

[0010] a first analysis module configured to obtain an average electrode inconsistency index of the first channel and the second channel according to the time series data based on a first model, the first model comprises: $JSD_t =$

$\frac{1}{n} \sum_{i=1}^m \left(\frac{1}{m} ICI(a_{it}) \right)$, wherein JSD represents a Jensen-Shannon (JS) divergence score function, t represents time, a_{it} represents the time series data of a corresponding channel of the first channel and the second channel recorded for the subject at a time point of t seconds, m represents the total number of a channel, n represents a total duration of the time series data of the corresponding channel, and ICI represents an inconsistency index of the corresponding channel;

[0011] a prediction module configured to compare the average electrode inconsistency index of the subject with a preset evaluation threshold, and predict that the subject is a patient with epilepsy when the average electrode inconsistency index of the subject is not lower than the preset evaluation threshold;

[0012] wherein the first channel is F3, C4, P4, O2, T7, FC1, CP1, CP6, TP9, TP10, F2, C1, C2, P2, FC3, FC4, CP3, CP4, PO4, P6, PO8, CPz, and POz, and the second channel is gastric body channel, lesser curvature channel, ascending colon channel, descending colon channel, and rectal channel.

[0013] In some embodiments, the system further comprises a second analysis module configured to convert the time series data into probability density distribution data.

[0014] In some embodiments, the second analysis module is further configured to obtain the inconsistency index of the corresponding channel according to the probability density distribution data based on a second model, the second model comprises: $ICI(a_{it}) = \sum_{i=1}^n \left(\frac{1}{2} D_{KL} \left(P(a_{it}) \middle| \frac{P(a_{it})+Q(b_{it})}{2} \right) + \frac{1}{2} D_{KL} \left(Q(b_{it}) \middle| \frac{P(a_{it})+Q(b_{it})}{2} \right) \right)$, wherein ICI represents an inconsistency index function, D_{KL} represents a Kullback-Leibler (KL) divergence function, t represents time, a_{it} represents the time series data of the corresponding channel recorded for the subject at the time point of t seconds, b_{it} represents the time series data of the corresponding channel recorded for healthy population at the time point of t seconds, n is a total duration of the time series data of the corresponding channel, $P(a_{it})$ represents sample distribution, and $Q(b_{it})$ represents reference distribution.

[0015] In some embodiments, the system further comprises a database configured to store data obtained by the data acquisition module, the data preprocessing module, the first analysis module, the second analysis module, and the prediction module, and generate a corresponding sample data set.

[0016] In some embodiments, the system further comprises an evaluation threshold setting module configured to set the average electrode inconsistency index that corresponds to the minimum critical value as the evaluation threshold by randomly obtaining the average electrode inconsistency index of 50%-70% of samples of patients with epilepsy in the sample data set and calculating critical values of of epilepsy onset thereof.

[0017] In some embodiments, the EEG data and the EGEG data are acquired synchronously.

[0018] In some embodiments, an acquisition duration of the EEG data and the EGEG data is at least 5 minutes.

[0019] In some embodiments, the preprocessing comprises filtering the EEG data to remove EEG signal below 1 Hz and above 30 Hz, and normalizing the EGEG data and the filtered EEG data to obtain the time series data.

[0020] In some embodiments, the data is acquired before or after a meal.

[0021] In some embodiments, the data is acquired from the subject in a normal state. As used herein, the term “normal state” refers to a state in which the subject exhibits normal electroencephalographic activity, i.e., no epileptiform discharges are detectable in the subject’s electroencephalogram. In some embodiments, when the subject is a healthy individual (i.e., not a patient with epilepsy), the normal state refers to any state of the subject. In some embodiments, when the subject is a patient with epilepsy, the normal state refers to a period other than seizure episodes (i.e., the interictal period).

[0022] In some embodiments, the subject is a patient with epilepsy.

[0023] In some embodiments, the patient with epilepsy is of focal seizure or generalized seizure.

[0024] Compared with the prior art, the present invention at least has the following beneficial effects.

[0025] The present invention provides a system for predicting epilepsy based on synchronous acquisition of EEG and EGEg signals. In the prior art, the method for processing EEG and EGEg signals is to apply phase-amplitude coupling, phase-phase coupling, and amplitude-amplitude coupling to the EEG and EGEg signals in different frequency bands. In the prior art, there are no reports on the use of the EEG and EGEg signals for the prediction and diagnosis of epilepsy.

[0026] Different from the conventional approach of extracting frequency bands from an EEG signal, the present invention achieves accurate prediction of epilepsy by collecting EEG data and EGEg data corresponding to EEG channels and EGEg channels by processing means, such as collecting EEG data and EGEg data corresponding to simplified EEG channels and EGEg channels (i.e., first and second channels (also referred to as EEG channels and EGEg channels after reduction) in the present invention) and combining with the inconsistency index and other processing methods. It should be emphasized that, in an aspect, the present invention breaks through the general idea of simultaneously using channels in relatively high correlation from the EEG channel and the EGEg channel, and realizes the reduction of channels by removing the EEG channels in relatively high correlation with the EGEg channels and replacing them with particular EGEg channels, thereby reducing the actual operational difficulty while improving the accuracy and sensitivity of diagnosing epilepsy. In another aspect, the present

invention does not detect epileptic seizures in patients with epilepsy by detecting abnormal discharges in the EEG, but diagnoses whether the subject is a patient with epilepsy based on the EEG (and EGEG) of the subject in the normal state.

[0027] In large-scale screening, even an experienced clinical doctor will find it difficult to judge whether a subject is a patient with epilepsy under time constraints and in the absence of diagnostic evidence just by virtue of his or her verbal description of the symptoms of “epileptic” seizures. The prediction model obtained based on the simplified channels and particular algorithms in the present invention are relatively high in accuracy and sensitivity, can capture and amplify differences between patients with epilepsy who have almost the same EEG activity in conventional methods (such as visual observation) and healthy population, and further distinguish them. It is suitable for assisting in the diagnosis of a subject who may not have abnormal discharges in the EEG in large-scale screening (such as healthy population who have symptoms similar to epileptic seizures but are not patients with epilepsy, and patients with epilepsy who have infrequent epileptic seizures).

[0028] Based on this, the system provided by the present invention not only requires short data collection time, but also can more accurately and efficiently distinguish patients with epilepsy from healthy population, and can be applied as an auxiliary diagnosis tool to primary screening of patients with epilepsy in relatively large population (such as communities and physical examination centers).

DESCRIPTION OF DRAWINGS

[0029] To make the embodiments of the present invention or the technical solutions in the prior art clearer, the drawings required to be used in the description of the embodiments or the prior art will be briefly introduced below. In all drawings, similar elements or components are generally identified by similar reference numerals. In the drawings, each element or component may not be drawn in actual proportion. It is obvious that the drawings described below are some embodiments of the present invention, and that other drawings can be obtained from these drawings for those of ordinary skill in the art without making inventive effort.

[0030] FIG. 1 is a location diagram of a first channel and a second channel after reduction in the present invention.

[0031] FIG. 2 is a result diagram of performances of different prediction models obtained by the present invention.

[0032] FIG. 3 is a flow diagram of a method for analysis of synchronously acquired EEG and EGEG signals as provided by the present invention.

[0033] FIG. 4 is a diagram of a system for predicting epilepsy based on synchronous acquisition of EEG and EGEG signals as provided by the present invention.

[0034] 100 - prediction system, 102 - data acquisition module, 104 - data preprocessing module, 106 - first analysis module, 108 - second analysis module, 110 - prediction module, 112 - database, and 114 - evaluation threshold setting module.

DETAILED DESCRIPTION

[0035] To make the objective, the technical solutions and advantages of the embodiments of the present invention clearer, the technical solutions of the embodiments of the present invention will be clearly and completely described below in combination with drawings. It is obvious that the described embodiments are some of the embodiments of the present invention, not all of the embodiments. Based on the embodiments of the present invention, all other embodiments obtained by those of ordinary skill in the art without making inventive effort shall belong to the protection scope of the present invention.

[0036] As used herein, such suffixes for representing elements as “module”, “component”, or “unit” are only for the purpose of facilitating the description of the present invention and do not have particular meanings in themselves. Therefore, “module”, “component”, or “unit” can be used in a mixed manner.

[0037] As used herein, orientation or positional relationship indicated by the terms such as “up”, “down”, “inside”, “outside”, “front”, “rear”, “one end”, and “the other end” is based on the orientation or positional relationship shown in the figures, only for the convenience of describing the present invention and simplifying the description, rather than for indicating or implying that the device or element referred to must have a particular orientation or must be constructed and operated in a particular orientation. Therefore, these terms cannot be understood as limitations on the

present invention. In addition, the terms “first” and “second” are only used for descriptive purposes, and cannot be deemed to indicate or imply relative importance.

[0038] As used herein, unless otherwise specified and limited, the terms such as “mount”, “arrange”, and “connect” should be understood broadly. For example, “connect” can be a fixed connection, a detachable connection, or an integral connection; it can be a mechanical connection, a direct connection, an indirect connection through an intermediate medium, or an internal connection between two components. In accordance with specific conditions, persons skilled in the art can understand the specific meanings of the above terms in the present invention.

[0039] As used herein, the term “and/or” refers to any and all combinations of one or more of the associated listed items.

[0040] As used herein, “a plurality of” means two or more, including but not limited to two, three, four, five, and so on.

[0041] It should be noted that the term “include”, “comprise” or any variant thereof is intended to encompass nonexclusive inclusion so that a process, method, article or device including a series of elements includes not only those elements but also other elements which have been not listed definitely or an element(s) inherent to the process, method, article or device. Moreover, the expression “comprising a(n) . . . ” in which an element is defined will not preclude presence of an additional identical element(s) in a process, method, article or device comprising the defined element(s) unless further defined.

[0042] As used herein, the term “about”, typically means $\pm 5\%$ of the stated value, more typically $\pm 4\%$ of the stated value, more typically $\pm 3\%$ of the stated value, more typically, $\pm 2\%$ of the stated value, even more typically $\pm 1\%$ of the stated value, and even more typically $\pm 0.5\%$ of the stated value.

[0043] Throughout this application, embodiments of this invention may be presented with reference to a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of the invention. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well as individual numerical values within that range. For example, description of a range such as “from 1 to 6” should be considered to have specifically disclosed subranges such as “from 1 to 3”, “from 1 to 4”, “from 1 to 5”,

“from 2 to 4”, “from 2 to 6”, “from 3 to 6”, etc.; as well as individual numbers within that range, for example, 1, 2, 3, 4, 5, and 6. This applies regardless of the breadth of the range.

[0044] Embodiment 1

[0045] 1.1 Dataset

[0046] Subject: this embodiment includes 102 patients with epilepsy of focal seizure or generalized seizure (57 males, aged 33.82 ± 12.65 years) and 46 healthy controls (26 males, aged 31.00 ± 8.45 years).

[0047] 1.1.1 EEG signal recording

[0048] The subjects are seated in a quiet room in preparation for the test. EEG data in 5-minute resting state is acquired from all subjects by using 64 Ag/AgCl electrodes (Brain Products, Munich, Germany), which are placed after the 10/20 system that has an impedance of below 10 k Ω . The EEG data is amplified and acquired at a sampling rate of 1000 Hz by using an actiCAP amplifier (BrainVision system amplifier, Brain Products, Munich, Germany). The subjects are instructed to fix their gaze on the center of the screen and avoid body movements as much as possible during the EEG recording.

[0049] 1.1.2 EGEG signal recording

[0050] Gastrointestinal myoelectrical activity is measured by using an 8-channel electrogastroenterography device (XDJ-S8, Hefei Kaili Co., Hefei, China). The subjects are instructed to abstain from alcohol, spicy, or irritating food for at least three days and fast for more than 6 hours before the test. Measurements are performed in the supine position. Four gastric electrodes (reflecting the gastric body, gastric antrum, lesser gastric curvature, and greater gastric curvature) and four intestinal electrodes (reflecting the ascending colon, transverse colon, descending colon, and rectum) (Hanjie Co., Shanghai, China) are placed on the abdominal skin. The specific placement sites for the aforementioned electrodes are as follows. Gastric body: 3-5 cm to the left of the midpoint of the line connecting the xiphoid process and the umbilicus, and 1 cm superiorly; gastric antrum: 2-4 cm to the right of the midpoint of the line connecting the xiphoid process and the umbilicus; lesser gastric curvature: halfway upward from the midpoint of the line connecting the xiphoid process and the umbilicus; greater gastric curvature: halfway downward from the midpoint of the line connecting the xiphoid process and the umbilicus. Ascending

colon: 2-4 cm to the right of a point level with the umbilicus; transverse colon: 1 cm below the umbilicus; descending colon: 2-4 cm to the left of a point level with the umbilicus; rectum: below the coccyx at the posterior lower region. A ground electrode is placed on the medial ankle of the right leg, and a reference electrode is placed on the medial wrist of the right hand. The subjects are instructed to remain quiet and avoid moving or talking during the test. In addition, after 6 minutes of preprandial EGEG recording, a meal-related functional load test is performed with about 200 kcal of food intake.

[0051] 1.2 Prediction algorithm based on Jensen-Shannon divergence (JSD)

[0052] 1.2.1 Data preprocessing

[0053] The dataset is divided into a training set and a testing set at a ratio of 7:3. For independent training, only EEG or EGEG signals are used. For EEG-EGEG multimodal fusion training, EEG and EGEG signals are used.

[0054] Firstly, EEGLAB in MATLAB is used to filter a raw EEG signal, and the data integrity is maintained by eliminating signals below 1Hz and above 30Hz. Secondly, the filtered EEG data and the raw EGEG signal are normalized. The feature matrix of the preprocessed EEG time series or the preprocessed EGEG time series is shown below.

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

[0055] Wherein m and n respectively represent dimensions of the feature matrix of the EEG and EGEG signals. m is the total number of electrodes, thus the i -th row represents the recording per second at the electrode e_i . n is the total duration of the EEG time series signal or the EGEG time series signal, i.e., the entire EEG or EGEG data comprises n seconds, thus the t -th column represents the EEG or EGEG data measured at each electrode at a time point of t seconds. Therefore, a_{it} is the EEG or EGEG signal of the electrode e_i recorded at the time point of t seconds.

[0056] 1.2.2 Conversion of preprocessed EEG time series or preprocessed EGEG time series into probability density distribution data

[0057] The data of each electrode is converted into probability density distribution data. In the training set, the time series data a_{it} of each electrode is used to fit the sample distribution $P(a_{it})$, wherein the mean μ_1 and the standard deviation σ_1 are used as sample features. In addition, for the healthy control data in the training set,

the time series data b_{it} of each electrode is fitted to the reference distribution $Q(b_{it})$, wherein the mean $u2_i$ and the standard deviation $\sigma2_i$ are used as discriminative reference features.

$$P(a_{it}) = \frac{1}{\sigma1_i\sqrt{2\pi}} e^{-\frac{(u1-u1_i)^2}{2\sigma1_i}}$$

[0058]

[0059] Wherein $P(a_{it})$ represents sample distribution, a_{it} represents the time series data of each electrode of a sample, $u1_i$ represents the mean of the sample, and $\sigma1_i$ represents the standard deviation of the sample.

$$Q(b_{it}) = \frac{1}{\sigma2_i\sqrt{2\pi}} e^{-\frac{(u2-u2_i)^2}{2\sigma2_i}}$$

[0060]

[0061] Wherein $Q(b_{it})$ represents reference distribution, b_{it} represents the time series data of each electrode for reference, $u2_i$ represents the mean for reference, and $\sigma2_i$ represents the standard deviation for reference.

[0062] When the difference between the sample features and the reference features is small, it means that there is no substantial difference between the sample distribution and the reference distribution, thereby indicating the absence of epilepsy. On the contrary, when the difference is considerable, it means that there is a significant difference between the sample distribution and the reference distribution, thereby indicating the presence of epilepsy.

[0063] 1.2.3 Construction of inconsistency indicators based on JSD

[0064] JS divergence is used to quantify the deviation between the sample distribution and the reference distribution, so as to measure the difference between different states (i.e., epilepsy and healthy controls); the inconsistency index (ICI) of each electrode is calculated to represent the difference between time points.

$$D_{KL} \left(P(a_{it}) \middle| \frac{P(a_{it})+Q(b_{it})}{2} \right) = \sum_{t=1}^T P(a_{it}) \log \frac{P(a_{it})}{\frac{1}{2}P(a_{it})+\frac{1}{2}Q(b_{it})}$$

[0065]

[0066] Wherein D_{KL} represents a KL divergence function, $P(a_{it})$ represents sample distribution, $Q(b_{it})$ represents reference distribution, a_{it} represents the time series data of a corresponding channel of a sample recorded at the time point of t seconds, b_{it} represents the time series data of the corresponding channel for

reference recorded at the time point of t seconds, t represents time, and T represents the total time.

$$ICI(a_{it}) = \sum_{i=1}^n \left(\frac{1}{2} D_{KL}(P(a_{it}) \parallel \frac{P(a_{it})+Q(b_{it})}{2}) + \frac{1}{2} D_{KL}(Q(b_{it}) \parallel \frac{P(a_{it})+Q(b_{it})}{2}) \right)$$

[0067]

[0068] ICI represents an inconsistency index function, $P(a_{it})$ represents sample distribution, $Q(b_{it})$ represents reference distribution, a_{it} represents the time series data of a corresponding channel of a sample recorded at the time point of t seconds, b_{it} represents the time series data of the corresponding channel for reference recorded at the time point of t seconds, t represents time, and n is the total duration of the time series data of the corresponding channel.

[0069] By calculating ICI , the fluctuations at all time points caused by the reference distribution are quantified. The average ICI score of each electrode can be used as a quantitative measure to evaluate the degree of distribution differences over time. When the average ICI score is relatively low, it indicates that there is no substantial difference between the sample distribution of the electrode and the reference distribution of the electrode. On the contrary, when the average ICI score is relatively high, it indicates a significant difference between the sample distribution of the electrode and the reference distribution of the electrode.

[0070] 1.2.4 Identification of DNB electrodes

[0071] Based on the DNB (Dynamic Network Biomarker) theory, it is assumed that a set of known molecules that emit pre-disease state signals of complex diseases prior to significant transitions are used to calculate the average ICI score of each electrode. Subsequently, the above scores are ranked according to the relative importance of key moments. The fluctuations caused by the reference distribution between electrodes at all time points are obtained according to the above equation, and then the ICI scores of the electrodes are ranked.

[0072] In this embodiment, the selection of DNB electrodes is divided into two different strategies. For the independent training of EEG, the electrodes ranked in the top 50% of the highest average scores are classified as DNB electrodes.

[0073] For EEG-EGEG multimodal fusion training, a total of 78 channels (62 EEG channels and 16 EGEG channels) are included. Firstly, the top 30 channels with the highest average score of individuals are selected. Then, if more than half of the individuals in the training set appeared in a certain channel, they would be used as

DNB electrodes. Since different electrodes may be associated with different regions of the brain or stomach, the higher the score of a particular electrode in the entire array, the more important its relationship with the epileptic activity.

[0074] 1.2.5 Calculation of JSD of global time series

[0075] The average electrode inconsistency index of the global time series is defined as follows:

[0076]
$$JSD_t = \frac{1}{n} \sum_{i=1}^m \left(\frac{1}{m} ICI(a_{it}) \right).$$

[0077] Wherein JSD represents a JS divergence score function, t represents time, a_{it} represents the time series data of the corresponding channel recorded at the time point of t seconds, m represents the total number of channels, n represents the total duration of the time series data of the corresponding channel, and ICI represents the inconsistency index of the corresponding channel.

[0078] The JSD score is used to quantify the distribution differences between all the electrodes, thereby helping to identify any significant deviations between distributions. The higher the score, the more significant the difference between distributions, and when the difference is exceeded, it indicates that the difference has escalated to a critical level, thereby predicting epilepsy. To find the appropriate threshold level for each patient, the following training scenario can be used: using 70% of epilepsy data as training data to obtain a threshold. For each epilepsy case in the training dataset, according to the JSD score, the critical values in the DNB theory for each onset can be obtained, and the minimum value of these critical values can be used as the threshold. When the score at a certain time point exceeds the threshold, the subject will be diagnosed with epilepsy.

[0079] 1.3 Correlation analysis between EEG channels and EGEG channels

[0080] According to the following equation, the probability distribution of each channel is evaluated by analyzing the EEG data of 62 channels and the EGEG data of 16 channels of each individual:

[0081]
$$P(a_{it}) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(u-u_i)^2}{2\sigma_i^2}}$$

[0082] The P-value is calculated by using the density function $P(a_{it})$ obtained by the EEG and EGEG channels and the average value and the standard deviation of

$P(a_{it})$. The correlation between the EEG and EGEG channels is determined by comparing the P-values thereof. A P-value less than 0.05 indicated no significant difference between the two, thus supporting the hypothesis of a relationship therebetween. To identify the EEG channels in the strongest correlation with the EGEG channels, the top three EEG channels with the smallest statistical deviation from the distribution of 16 EGEG channels in each sample are selected.

[0083] Embodiment 2

[0084] It should be noted that in the present invention, the selection of DNB electrodes is carried out in two different ways. For independent EEG training, the top 50% of electrodes with the highest average ICI score are selected as DNB electrodes based on the total number of electrodes as used. For EEG-EGEG multimodal fusion training, the top 30 channels with the highest average ICI score of each individual in the training set are selected. If a channel appears in more than half of the training set, it is designated as a common DNB electrode for predicting epilepsy. As the results are shown in FIG. 1, the 30 channels obtained by the present invention include: 23 EEG channels (first channel): F3, C4, P4, O2, T7, FC1, CP1, CP6, TP9, TP10, F2, C1, C2, P2, FC3, FC4, CP3, CP4, PO4, P6, PO8, CPz, and POz; 7 EGEG channels (second channel): gastric body channel (i.e., electrode 1), lesser gastric curvature channel (i.e., electrode 2), gastric antrum (i.e., electrode 4), ascending colon channel (i.e., electrode 5), transverse colon channel (i.e., electrode 6), descending colon channel (i.e., electrode 7), and rectal channel (i.e., electrode 8).

[0085] In FIG. 2, EEG represents a prediction model obtained from the training with data acquired from 62 EEG channels in the method of the present invention; EEG* represents a prediction model obtained from the training with data acquired from the top 50% of electrodes with the highest average ICI scores among 62 EEG channels in the method of the present invention; EEG+EGEG represents a prediction model obtained from the training with data acquired from 78 channels (i.e., 62 EEG channels and 16 EGEG channels) in the method of the present invention; EEG+EGEG* represents a prediction model obtained from the training with data acquired from the first and second channels in the method of the present invention.

[0086] As shown in FIG. 2, the prediction models obtained by processing EEG or EEG+EGEG data in the method of the present invention have shown relatively good

predictive ability. However, while the prediction models need to collect and calculate more data, the accuracy thereof can hardly reach 95%. In the prediction model constructed on the basis of the electrodes obtained by reduction in the present invention, there is a decrease in the number of the electrodes that need to be placed, and yet the performance in predicting epilepsy is improved, with an accuracy of 98.92% and a sensitivity of 98.82% observed.

[0087] Correlation analysis between EEG channels and EGEG channels.

[0088] This embodiment further explores whether there is a certain correlation between different EEG channels and EGEG channels in predicting epilepsy through probability distribution. Specifically, this embodiment analyzes the probability distribution of 62 EEG channels and 16 EGEG channels for each subject. In the 148 samples, the frequency of occurrence of 62 EEG channels is statistically similar to that of 16 EGEG channels. Further, this embodiment extracts the top three EEG channels correlated to each EGEG channel in the subjects, and these channels are defined as key correlation factors of particular EGEG channels.

[0089] A similarity comparison is made between the distribution of all the EGEG channels and that of all the EEG channels, and the number of occurrences of eight EGEG leads that does not show significant differences from the distribution of EEG channels in 148 subjects is calculated. The more occurrences, the higher the correlation between the EEG leads and the EGEG. This embodiment determines that the top 20% of EEG channels do not show statistically significant distribution differences compared with 16 EGEG channels, and designates them as highly correlated EEG channels, as shown in Table 1.

[0090] Table 1: statistical analysis of the frequency of occurrence of the top 20% of EEG channels correlated to EGEG in 148 sample results.

Name	AF8	Fp1	T8	AF7	Fp2	T7	FT10	Fpz	TP9	P8	F8	AF4
Electrode	53	1	14	52	2	13	29	60	30	16	12	39
Frequency	78	49	49	48	43	41	40	40	39	38	37	36

[0091] The results show that most of the EEG channels shown in Table 1 are not included in the first channel. Based on this, the present invention believes that the function of most of the EEG channels shown in Table 1 in predicting epilepsy can be replaced by correlated EGEG channels (i.e., the second channel as mentioned above).

[0092] Embodiment 3

[0093] Referring to FIG. 3, the present invention provides a method for analysis of synchronously acquired EEG and EGEG signals, comprising the following steps:

[0094] S101. inputting synchronously acquired EEG data and EGEG data of a subject;

[0095] In some embodiments, the EEG data is acquired from a first channel, and the EGEG data is acquired from a second channel.

[0096] In the present invention, “channel”, “electrode”, and “lead” have the same meaning, unless otherwise specified.

[0097] In some embodiments, the first channel comprises F3, C4, P4, O2, T7, FC1, CP1, CP6, TP9, TP10, F2, C1, C2, P2, FC3, FC4, CP3, CP4, PO4, P6, PO8, CPz, and POz; the second channel comprises gastric body channel, lesser gastric curvature channel, gastric antrum channel, ascending colon channel, transverse colon channel, descending colon channel, and rectal channel.

[0098] It should be noted that the first channel and the second channel are respectively the EEG channels and EGEG channels after reduction in the present invention, and the correlation between the first channel and the second channel in terms of epilepsy is relatively low.

[0099] Different from the prior art of coupling an EGG signal with an EEG signal, the present invention obtains EEG channels in relatively strong correlation with EGEG in patients with epilepsy, but excludes these EEG channels from the EEG channels to be acquired in practice, thereby obtaining the first channel as mentioned above. The present invention has found that these EEG channels can be replaced with correlated EGEG channels (i.e., the second channel as mentioned above) in terms of predictive function, and the adoption of the first and second channels as mentioned above, which have relatively low correlation in predicting epilepsy, not only reduces the difficulty of acquisition operations, but also improves the accuracy of predicting patients with epilepsy, making it more convenient in practical applications. However, the employment of a smaller number of EEG channels (such as conventional 16 leads and 32 leads) is not sufficient to cover the entire brain, making it difficult to predict epilepsy.

[0100] In some embodiments, an acquisition duration of the EEG data and the EGEG data is at least 5 minutes.

- [0101]** In some embodiments, the data is acquired before or after a meal.
- [0102]** In some embodiments, the data is acquired from the subject in a normal state.
- [0103]** In some embodiments, the subject is a patient with epilepsy.
- [0104]** In some embodiments, the patient with epilepsy is of focal seizure or generalized seizure.
- [0105]** S102. preprocessing the synchronously acquired EEG data and EGEG data to obtain corresponding time series data;
- [0106]** In some embodiments, the preprocessing comprises filtering the EEG data to remove EEG signal below 1Hz and above 30Hz, and normalizing the EGEG data and the filtered EEG data to obtain the time series data.
- [0107]** S103. obtaining an average electrode inconsistency index of the first channel and the second channel based on a first model;
- [0108]** In some embodiments, the first model comprises: $JSD_t = \frac{1}{n} \sum_{i=1}^m \left(\frac{1}{m} ICI(a_{it}) \right)$, wherein JSD represents a JS divergence score function, t represents time, a_{it} represents the time series data of a corresponding channel of the first channel and the second channel recorded for the subject at a time point of t seconds, m represents the total number of a channel, n represents a total duration of the time series data of the corresponding channel, and ICI represents an inconsistency index of the corresponding channel.
- [0109]** In some embodiments, the corresponding channel refers to F3, C4, P4, O2, T7, FC1, CP1, CP6, TP9, TP10, F2, C1, C2, P2, FC3, FC4, CP3, CP4, PO4, P6, PO8, CPz, POz, gastric body channel, lesser gastric curvature channel, gastric antrum channel, ascending colon channel, transverse colon channel, descending colon channel, and rectal channel. In some embodiments, the total duration refers to the total duration of the entire EEG data and EGEG data that are acquired synchronously. For example, when the acquisition duration is 5 minutes, the total duration is 5 minutes.
- [0110]** Based on the first and second channels as mentioned above, the present invention not only reduces the difficulty of acquisition operations, but also reduces the amount of data that needs to be analyzed and calculated, thereby improving the

computation efficiency, while maintaining relatively high levels of accuracy and sensitivity.

[0111] In some embodiments, the method further comprises: S201. converting the time series data into probability density distribution data.

[0112] In some embodiments, the method further comprises: S202. obtaining the inconsistency index of the corresponding channel according to the probability density distribution data based on a second model, the second model comprises: $ICI(a_{it}) = \sum_{i=1}^n (\frac{1}{2} D_{KL}(P(a_{it}) | \frac{P(a_{it})+Q(b_{it})}{2}) + \frac{1}{2} D_{KL}(Q(b_{it}) | \frac{P(a_{it})+Q(b_{it})}{2}))$, wherein ICI represents an inconsistency index function, D_{KL} represents a KL divergence function, t represents time, a_{it} represents the time series data of the corresponding channel recorded for the subject at the time point of t seconds, b_{it} represents the time series data of the corresponding channel recorded for a healthy population at the time point of t seconds, n is a total duration of the time series data of the corresponding channels, $P(a_{it})$ represents sample distribution, and $Q(b_{it})$ represents reference distribution.

[0113] S104. comparing the average electrode inconsistency index of the subject with a preset evaluation threshold, and predicting that the subject is a patient with epilepsy when the average electrode inconsistency index of the subject is not lower than the preset evaluation threshold.

[0114] In some embodiments, the method for obtaining the preset evaluation threshold comprises: S203. randomly obtaining the average electrode inconsistency index of 50%-70% of samples of patients with epilepsy in the sample data set and calculating critical values of epilepsy onset (based on the DNB theory), and setting the average electrode inconsistency index that corresponds to the minimum critical value as the evaluation threshold.

[0115] Embodiment 4

[0116] Referring to FIG. 4, the present invention provides a system (prediction system) 100 for predicting epilepsy based on synchronous acquisition of EEG and EGEG signals, comprising:

[0117] a data acquisition module 102 configured to obtain data comprising EEG data and EGEG data;

- [0118]** In some embodiments, the EEG data and the EGEG data are acquired synchronously.
- [0119]** In some embodiments, an acquisition duration of the EEG data and the EGEG data is at least 5 minutes.
- [0120]** In some embodiments, the data is acquired before or after a meal.
- [0121]** In some embodiments, the EEG data is acquired from a first channel, and the EGEG data is acquired from a second channel.
- [0122]** In some embodiments, the data comprises sample data from sample population and subject data from a subject.
- [0123]** In some embodiments, the sample population comprises healthy population and patients with epilepsy.
- [0124]** In some embodiments, the data is acquired from the subject in a normal state.
- [0125]** In some embodiments, the subject is a patient with epilepsy.
- [0126]** In some embodiments, the patient with epilepsy is of focal seizure or generalized seizure.
- [0127]** In some embodiments, the first channel comprises F3, C4, P4, O2, T7, FC1, CP1, CP6, TP9, TP10, F2, C1, C2, P2, FC3, FC4, CP3, CP4, PO4, P6, PO8, CPz, and POz; the second channel comprises gastric body channel, lesser gastric curvature channel, gastric antrum channel, ascending colon channel, transverse colon channel, descending colon channel, and rectal channel.
- [0128]** a data preprocessing module 104 configured to preprocess the data to obtain corresponding time series data;
- [0129]** In some embodiments, the preprocessing comprises filtering the EEG data to remove EEG signal below 1Hz and above 30Hz, and normalizing the EGEG data and the filtered EEG data to obtain the time series data.
- [0130]** a first analysis module 106 configured to obtain an average electrode inconsistency index of the first channel and the second channel based on a first model according to the time series data;
- [0131]** In some embodiments, the first model comprises: $JSD_t = \frac{1}{n} \sum_{i=1}^m \left(\frac{1}{m} ICI(a_{it}) \right)$, wherein JSD represents a JS divergence score function, t represents time, a_{it} represents the time series data of a corresponding channel of

the first channel and the second channel recorded for the subject at a time point of t seconds, m represents the total number of a channel, n represents a total duration of the time series data of the corresponding channel, and ICI represents an inconsistency index of the corresponding channel.

[0132] a second analysis module 108 configured to convert the time series data into probability density distribution data.

[0133] In some embodiments, the second analysis module 108 is further configured to obtain the inconsistency index of the corresponding channel based on a second model according to the probability density distribution data.

[0134] In some embodiments, the second model comprises: $ICI(a_{it}) = \sum_{i=1}^n (\frac{1}{2} D_{KL}(P(a_{it}) | \frac{P(a_{it})+Q(b_{it})}{2}) + \frac{1}{2} D_{KL}(Q(b_{it}) | \frac{P(a_{it})+Q(b_{it})}{2}))$, wherein ICI represents an inconsistency index function, D_{KL} represents a KL divergence function, t represents time, a_{it} represents the time series data of the corresponding channel recorded for the subject at the time point of t seconds, b_{it} represents the time series data of the corresponding channel recorded for healthy population at the time point of t seconds, n is a total duration of the time series data of the corresponding channels, $P(a_{it})$ represents sample distribution, and $Q(b_{it})$ represents reference distribution.

[0135] In some embodiments, the first analysis module 106 and the second analysis module 108 can be merged into an analysis module.

[0136] a prediction module 110 configured to compare the average electrode inconsistency index of the subject with a preset evaluation threshold, and predict that the subject is a patient with epilepsy when the average electrode inconsistency index of the subject is not lower than the preset evaluation threshold;

[0137] a database 112 configured to store the data obtained by the data acquisition module 102, the data preprocessing module 104, the first analysis module 106, the second analysis module 108, and the prediction module 110, and generate a corresponding sample data set.

[0138] In some embodiments, the sample data set comprises a data set of healthy population and a data set of patients with epilepsy.

[0139] an evaluation threshold setting module 114 configured to set the average electrode inconsistency index that corresponds to the minimum critical value as the

evaluation threshold by randomly obtaining the average electrode inconsistency index of 50%-70% of samples of patients with epilepsy in the sample data set and calculating critical values of epilepsy onset thereof.

[0140] Through the description of the above manners of implementation, persons skilled in the art can clearly understand that the above embodiment methods can be implemented through software and necessary general hardware platforms. Of course, they can also be done through hardware, but in many cases, the former is the preferred manner of implementation. Based on this understanding, the essence of the technical solutions of the present invention, or the part that contributes to the prior art, can be embodied in the form of a software product, which is stored in a storage medium (such as ROM/RAM, magnetic disk, optical disk) and comprises several instructions to enable a computer terminal (which can be a mobile phone, a computer, a server, or a network device, etc.) to perform the methods described in the embodiments of the present invention.

[0141] The embodiments of the present invention are described above with reference to the accompanying drawings, but the present invention is not limited to the aforementioned specific embodiments. The aforementioned embodiments are merely illustrative and not limiting. For those of ordinary skill in the art, many forms can be made under the teaching of present invention without departing from the spirit of the present invention and the scope of the claims, all of which shall fall within the protection scope of the present invention.

CLAIMS

Claim 1. A system for predicting epilepsy based on synchronous acquisition of electroencephalogram (EEG) and electrogastroenterogram (EGEG) signals, comprising:

a data acquisition module configured to acquire data comprising EEG data and EGEG data, wherein the data is acquired before or after a meal from a subject in a normal state, the EEG data is acquired from a first channel, the EGEG data is acquired from a second channel, the data comprises sample data from sample population and subject data from a subject, and the EEG data and the EGEG data are acquired synchronously;

a data preprocessing module configured to preprocess the data to obtain corresponding time series data; wherein a feature matrix of the time series data is represented as:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

wherein m and n respectively represent a dimension of the feature matrix of an EEG time series signal or an EGEG time series signal, m is the total number of an electrode, and n is a total duration of the EEG time series signal or the EGEG time series signal;

a first analysis module configured to obtain an average electrode inconsistency index of the first channel and the second channel according to the time series data based on a first model, the first model comprises: $JSD_t = \frac{1}{n} \sum_{i=1}^m \left(\frac{1}{m} ICI(a_{it}) \right)$,

wherein JSD represents a Jensen-Shannon divergence score function, t represents time, a_{it} represents the time series data of a corresponding channel of the first channel and the second channel recorded for the subject at a time point of t seconds, m represents the total number of channels, n represents the total duration of the time series data of the corresponding channel, and ICI represents an inconsistency index of the corresponding channel;

a prediction module configured to compare the average electrode inconsistency index of the subject with a preset evaluation threshold, and predict that the

subject is a patient with epilepsy when the average electrode inconsistency index of the subject is not lower than the preset evaluation threshold;
 wherein the first channel is F3, C4, P4, O2, T7, FC1, CP1, CP6, TP9, TP10, F2, C1, C2, P2, FC3, FC4, CP3, CP4, PO4, P6, PO8, CPz, and POz, and the second channel is gastric body channel, lesser gastric curvature channel, ascending colon channel, descending colon channel, and rectal channel.

Claim 2. The system according to claim 1, further comprising a second analysis module configured to convert the time series data into probability density distribution data.

Claim 3. The system according to claim 2, wherein the second analysis module is further configured to obtain the inconsistency index of the corresponding channel according to the probability density distribution data based on a second model, the second model comprises: $ICI(a_{it}) = \sum_{i=1}^n (\frac{1}{2} D_{KL}(P(a_{it}) | \frac{P(a_{it})+Q(b_{it})}{2}) + \frac{1}{2} D_{KL}(Q(b_{it}) | \frac{P(a_{it})+Q(b_{it})}{2}))$, wherein ICI represents an inconsistency index function, D_{KL} represents a Kullback-Leibler divergence function, t represents time, a_{it} represents the time series data of the corresponding channel recorded for the subject at the time point of t seconds, b_{it} represents the time series data of the corresponding channel recorded for healthy population at the time point of t seconds, n is a total duration of the time series data of the corresponding channel, $P(a_{it})$ represents sample distribution, and $Q(b_{it})$ represents reference distribution.

Claim 4. The system according to claim 2, further comprising a database configured to store data obtained by the data acquisition module, the data preprocessing module, the first analysis module, the second analysis module, and the prediction module, and generate a corresponding sample data set.

Claim 5. The system according to claim 4, further comprising an evaluation threshold setting module configured to set the average electrode inconsistency index that corresponds to the minimum critical value as the evaluation threshold by

randomly obtaining the average electrode inconsistency index of 50%-70% of samples of patients with epilepsy in the sample data set and calculating critical values of epilepsy onset thereof.

Claim 6. The system according to claim 1, wherein an acquisition duration of the EEG data and the EGEG data is at least 5 minutes.

Claim 7. The system according to claim 1, wherein the preprocessing comprises filtering the EEG data to remove EEG signal below 1 Hz and above 30 Hz, and normalizing the EGEG data and the filtered EEG data to obtain the time series data.