Computers and Electrical Engineering 68 (2018) 437–446

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Using methods of time series data mining to recognize the influences of environmental factors on epileptic seizures☆

Jhe-Nan Lin[a], Fu-Yi Yang[b], Wei-Zen Sun[c,d], Meng-Han Yang[a,*]

[a] Department of Computer Science and Information Engineering, National Kaohsiung University of Applied Sciences, No.415, Jiangong Rd., Sanmin Dist., Kaohsiung 807, Taiwan, ROC
[b] Department of Neurology, Taipei Tzu Chi General Hospital, No. 289, Jianguo Rd., Xindian District, New Taipei 23142, Taiwan, ROC
[c] Department of Anesthesiology, National Taiwan University Hospital, No.1, Changde St., Zhongzheng Dist., Taipei 10048, Taiwan, ROC
[d] Research Center for Emergency Medical Service, College of Medicine, National Taiwan University, No.1, Jen Ai Road Section 1, Taipei 10051, Taiwan, ROC

ARTICLE INFO

Keywords:
Correlation measure
EEMD
Environmental factor
Epileptic seizure
FGD
Statistical test

ABSTRACT

This research investigated whether environmental factors are important triggers for epileptic seizures. Epilepsy is a very common neurological disease whose exact pathology remains unknown. Numerous studies have observed that environmental changes can be important triggers for various illnesses. Incorporating a nationwide dataset, environmental factors were characterized as inducers for epilepsy using conventional methods and a recently-proposed algorithm of time series data mining. From iterative verification using these algorithms, epilepsy incidence increments were established in January and February of each year. Furthermore, epilepsy incidence was positively correlated with atmospheric pressure, and negatively correlated with ambient temperature. In summary, environmental factors might be important triggers for epilepsy, and the mined epidemiological properties are expected to be helpful for management of this disease.

1. Introduction

An epileptic seizure is a brief symptomatic episode caused by abnormal excessive or synchronous neuronal activity in the brain [1]. Epilepsy is a chronic non-communicable disorder that affects people of all ages. Approximately 50 million people worldwide have epilepsy, making it one of the most common neurological diseases globally [2]. Seizures include focal and generalized types whose signs and symptoms differ. Epilepsy can have both genetic and acquired causes, along with interaction of these factors in many cases [3]. However, the exact pathology of epilepsy is unknown. Epilepsy may occur as a result of other conditions including tumors, strokes, head trauma, previous infections of the central nervous system, and brain damage near birth [2,4]. Further investigation is needed to understand how the brain incurs epileptic conditions.

A time series is a collection of observations made chronologically, and is characterized by a numerical and continuous nature. The purpose of time series data mining is to discover hidden information or knowledge from either the original or the transformed time series data. During the last decade, several studies have focused on the temporal rhythms of disease [5–9]. As well, the impact of environmental factors on human physiological functions has been a key focus of scientific research. However, conflicting results have emerged on the impact of the geomagnetic field on epileptic seizures [10,11]. Likewise, solar beam intensity and shadow flicker

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. J-S Sheu.

* Corresponding author.
E-mail address: menghanyang@cc.kuas.edu.tw (M.-H. Yang).

https://doi.org/10.1016/j.compeleceng.2018.04.020
Received 14 September 2017; Received in revised form 24 April 2018; Accepted 25 April 2018

frequency, which stimulate retinal cells, are believed to cause epileptic seizures in susceptible people [12].

A previous study concluded that various meteorological factors are positively associated with the incidence of eclampsia [13]. However, the source data were limited and controversial conclusions about the impact of environmental factors on epileptic seizures were drawn [14–16]. Therefore, using a nationwide population-based dataset with statistical support, and techniques of time series data mining, the seasonal incidence patterns of epileptic seizures in Taiwan are presented. In addition to traditional statistical methods, the newly proposed Fourier-Gaussian Decomposition (FGD) algorithm [17], as well as the conventional ensemble empirical mode decomposition (EEMD), are employed to iteratively mine the temporal rhythms of epilepsy. Furthermore, correlation coefficients are used to measure the relationships between meteorological factors and epileptic seizures. The mined seasonal incidence patterns of epilepsy and their associations with various weather data may be helpful to better understand the characteristics of epileptic seizures, and would be beneficial in supporting appropriate management of this disease.

The reminder of this paper is organized as follows. Section 2 describes materials and methods used. Section 3 provides the evaluation outcomes and Section 4 discusses the analysis results. Finally, Section 5 summarizes conclusions.

2. Materials and methods

In time series data mining, statistical testing methods are used to validate the reliability of source data. The FGD algorithm and EEMD method were then utilized to explore the temporal rhythms of epileptic seizure. Uncovered seasonal patterns were compared and validated using statistical tests. Finally, correlation coefficients were employed to evaluate the associations between various meteorological factors and epilepsy.

2.1. The data sources of study

The database used in this study was acquired from the emergency medical service system, which is managed by Ministry of Health and Welfare, R.O.C. Raw data in the database came from ambulance paper sheets recorded nationwide by emergency medical technicians. This database includes numerous attributes, such as the date of rescue; patient age, gender, and medical history; the reason for calling the ambulance service; and the address of the accident site. Strict rules exist for the education and training of emergency medical technicians, and standard operation procedures for emergency treatment are well formulated. This provides confidence in the reliability of this database. The version of the emergency medical service database employed in this study (IRB number 201307056RIND) contains ambulance paper sheets recorded from 2007 to 2013.

The seasonal incidence patterns of epileptic seizures in Taiwan were analyzed, with the study group consisting of cases whose reasons for calling the ambulance service were coded as A120, which represents the occurrence of epileptic seizure. For each of the 84 months from January 2007 to December 2013, counts of monthly cases for epileptic seizures were recorded as observation values.

The seasonal incidence patterns of diseases may not be reflected directly by months because of geographical differences. Therefore, meteorological data were also adopted in this study to investigate their relationships with the incidence of epileptic seizures. Measured values of meteorological variables are monitored, maintained, and organized by local stations throughout Taiwan for the Central Weather Bureau, R.O.C. The observation data can be freely downloaded from the open platform for government information. Land-based meteorological information consists of the following items: accumulated precipitation; temperature; maximum temperature; minimum temperature; relative humidity; maximum relative humidity; minimum relative humidity; mean wind speed/direction; station pressure; maximum station pressure; minimum station pressure; and sunshine duration. For each of the meteorological data types, observations from the 84-month period between January 2007 and December 2013 were adopted in this study.

Before performing any analysis, these temporal sequences of source data would be scaled in order to removing impacts of different measurement units.

2.2. The statistical analysis

Statistical testing methods, such as the Kolmogorov–Smirnov, Ljung–Box, and Shapiro–Wilk tests were used to verify whether observation values corresponded to Gaussian white noise. The standard score (or z-score) was used to mine the seasonal incidence patterns of epileptic seizure, and goodness-of-fit was used to validate the statistical significance of the incidence patterns. Finally, correlation coefficients were adopted to investigate the association between meteorological variables and epileptic seizures. For these association analyses, partial correlation measures identified the interactions of different meteorological variables under consideration. Please refer to a previous report for detailed descriptions of these statistical methods [18].

2.3. The ensemble empirical mode decomposition (EEMD) method

Recently wavelet transforms have become popular in time series data mining. However, choosing a suitable wavelet for a problem typically relies on a priori assumptions. The Hilbert-Huang transform (HHT) has been proposed to generate intrinsic mode functions (IMFs), which can be empirically derived from data and represent hidden patterns in the time series [19]. HHT consists of the empirical mode decomposition (EMD) and Hilbert spectral analysis (HSA). This method uses the EMD to decompose a sequence of observations into a finite number of IMFs, then uses the HSA to obtain instantaneous frequencies. To overcome the scale mixing problem, the ensemble EMD (EEMD) has been further proposed. EEMD defines the true IMF components as the mean of an ensemble