**Use of Information Technologies to Explore Correlations between Climatic Factors and Spontaneous Intracerebral Hemorrhage in Different Age Groups**

**Hsien-Wei Ting**  
Department of Neurosurgery, Taipei Hospital, Ministry of Health and Welfare, New Taipei City, Taiwan;  
Department of Computer Science and Engineering, Yuan Ze University, Taoyuan City, Taiwan  
ting.ns@gmail.com

**Chien-Lung Chan**  
Department of Information Management, Yuan Ze University, Taoyuan City, Taiwan;  
Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Taoyuan City, Taiwan  
elchan@saturn.yzu.edu.tw

**Ren-Hao Pan**  
Department of Computer Science and Engineering, Yuan Ze University, Taoyuan City, Taiwan;  
Department of Information Management, Tunghai University, Taichung City, Taiwan  
pan@51donate.com

**Robert K. Lai**  
Department of Computer Science and Engineering, Yuan Ze University, Taoyuan City, Taiwan  
klrai@saturn.yzu.edu.tw

**Ting-Ying Chien***  
Department of Computer Science and Engineering, Yuan Ze University, Taoyuan City, Taiwan;  
Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Taoyuan City, Taiwan  
tinin@saturn.yzu.edu.tw

**Abstract**  
Spontaneous intracerebral hemorrhage (sICH) has a high mortality rate. Research has demonstrated that sICH occurrence is related to weather conditions; therefore, this study used the decision tree method to explore the impact of climatic risk factors on sICH at different ages. The Taiwan National Health Insurance Research Database (NHIRD) and other open-access data were used in this study. The inclusion criterion was a first-attack sICH. The decision tree algorithm and random forest were implemented in R programming language. We defined a high risk of sICH as more than the average number of cases daily, and the younger, middle-aged and older groups were calculated as having 0.77, 2.26 and 2.60 cases per day, respectively. In total, 22,684 sICH cases were included in this study; 3,102 patients were younger (<44 years, younger group), 9,089 were middle-aged (45–64 years, middle group), and 10,457 were older (>65 years, older group). The risk of sICH in the younger group was not correlated with temperature, wind speed or humidity. The

Published by **Transdisciplinary Journal of Computing Science and Engineering**  
Volume 11, Issue 4, December 2017, pp. 142-151  
Special Issue

Open Access  
http://dx.doi.org/10.5626/JCSE.2017.11.4.142  
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 25 May 2017; Accepted 25 November 2017  
*Corresponding Author

Copyright © 2017. The Korean Institute of Information Scientists and Engineers
middle group had two decision nodes: a higher risk if the maximum temperature was >19°C (probability = 63.7%), and if the maximum temperature was <19°C in addition to a wind speed <2.788 (m/s) (probability = 60.9%). The older group had a higher risk if the average temperature was >23.933°C (probability = 60.7%). This study demonstrated that the sICH incidence in the younger patients was not significantly correlated with weather factors; that in the middle-aged sICH patients was highly-correlated with the apparent temperature; and that in the older sICH patients was highly-correlated with the mean ambient temperature. “Warm” cold ambient temperatures resulted in a higher risk of sICH, especially in the older patients.

**Category:** Bioinformatics

**Keywords:** Big data; Climatic factors; Decision tree; Random forest; Spontaneous intracerebral hemorrhage

---

**I. INTRODUCTION**

Spontaneous intracerebral hemorrhage (sICH) accounts for 10%–35% of stroke patients. The incidence ranges from 10 to 60 cases per 100,000 population per year, with a high mortality rate [1, 2]. sICH patients require a high level of medical resources, especially in terms of emergency and intensive care facilities. Reduction of the incidence of sICH is a huge challenge in many countries. Research has demonstrated that sICH occurrence is related to weather conditions [3-5]; temperature factors have been discussed in depth in previous studies, and are considered important risk factors for sICH [3, 5, 6]. According to previously-reported evidence, the influences of climate risk factors vary in different age groups [5, 7]. However, previous studies have not explored the correlations between combinations of temperature factors and age groups. In other words, grouping by age is necessary to obtain a complete picture.

Researchers have long used information technology to improve medical decision-making [2]. Classification models for medical issues have been developed using artificial intelligence technologies [8, 9]. Previous studies have integrated patient data and climate data to identify correlations between diseases and climate risk factors [3-5, 10]. The decision tree method is one of the most important information technologies for prediction and diagnosis, by which meaningful and logical models that may be used for decision-making can be constructed [11, 12]. The decision tree method may also highlight meaningful factors that require further evaluation [13]. This technology is used in different fields for medical decision-making, and enables construction of models that are easy to understand [11].

In this study, we used the decision tree method, the C5.0 algorithm, the Classification and Regression Tree (CART) method, and the random forest method to explore the impacts of climatic risk factors on sICH at different ages based on the Taiwan National Health Insurance Research Database (NHIRD), which is a good longitudinal dataset for use in financial and epidemiological research. Several studies have used the NHIRD only for medical research [1, 2, 14, 15]. The dataset used in this study covered outpatient and inpatient claims data, with detailed information for each visit/stay longitudinally [16]. This study integrated the NHIRD and climate data obtained from the Taiwan Typhoon and Flood Research Institute and employed the decision tree method to identify and evaluate the correlations between weather factors and sICH in a Taiwanese population.

**II. METHODS**

**A. Data Source, Protection and Permission**

This study integrated the NHIRD, the climate database from the Data Bank for Atmospheric and Hydrologic Research, and the Government Open Data Platform in Taiwan with big data analytics systems using the platform of the Innovation Center for Big Data and Digital Convergence, Yuan Ze University [16, 17]. The personal information of all subjects was encrypted using a double scrambling protocol for research purposes to protect patient privacy. All researchers who wish to use the NHIRD and its data subsets are required to sign a written agreement declaring that they have no intention of obtaining information that could potentially violate the privacy of patients or care providers. This study was approved by the Institutional Review Board of Taipei Hospital (No. TH-IRB-0015-0003), and the protocol was evaluated by the National Health Research Institutes (NHRI), which consented to this planned analysis of the NHIRD (No. NHIRD-104-183).

**B. Data Management**

The inclusion criterion for patients in this research was a first-attack sICH, identified by a principal diagnosis code in the International Classification of Diseases 9th version (ICD-9) of 431. In total, 128,173 sICH cases were registered from 2001 to 2011 in Taiwan. Patients who were admitted due to traumatic intracranial hemorrhage (TICH) (ICD-9 codes 800.00 to 804.99, 850.00 to 854.19,
959.01, and 959.09) were excluded [1], which totaled 1,753 cases. A final study sample of 126,420 patients remained, consisting of first-attack sICH patients registered between 2001 and 2011 in Taiwan. We then extracted sICH patients of Taipei City, which left 22,648 subjects. Open-access data from the Data Bank for Atmospheric and Hydrologic Research and the Government Open Data Platform in Taiwan were also integrated in this study (Fig. 1).

C. Definitions of Climate Risk Factors

Data consisting of the daily temperature (°C), average wind speed (m/s) and average humidity (%) were collected from the climate data as climate risk factors for each time slot. This study defined different time slots for the average of the temperature (°C), the maximum temperature (°C), the minimum temperature (°C), the average wind speed (m/s) and average humidity (%); the average temperature ranges ran from 0 to 24 hours (time slot 1); 12 to 36 hours (time slot 2); 24 to 48 hours (time slot 3); 36 to 60 hours (time slot 4); 48 to 72 hours (time slot 5) and 60 to 84 hours (time slot 6) as features before the day of admission [3-5] (Fig. 2).

D. Data Analytics

Information on patient gender and age, date of admission and the climate risk factors in each time slot were collected. The data analytics algorithm C5.0, CART, and random forest were implemented in R programming language, which is a freeware and open-source code system [18]. The C5.0 package was implemented based on Quinlan’s C5.0 algorithm [19], which is a commercial package based on the C4.5 algorithm [11, 20-22]. The C5.0 algorithm is faster and requires less memory usage than C4.5. Using the C4.5 algorithm, decision trees are built using the divide-and-conquer approach and the concept of information entropy [23]. The input format for the C4.5 algorithm is an attribute-class dataset, which means that the object in the dataset consists of a target class and n-dimensional vectors that represent attribute values of the object. In this work, the cases per day of the group were used as the target class, the features of which are shown in Fig. 2. C4.5 chooses the attribute with the highest normalized
Use of ITs to Explore Correlations between Climatic Factors and sICH in Age Group

Hsien-Wei Ting et al. 145 http://jcse.kiise.org

information [22] gain as the split node, and then recurs on the subsets until the subset is pure or satisfies the stop criteria. As a result, the decision tree is built; the internal nodes of the tree represent the chosen attributes, and the left nodes show the proportion of each target class.

The CART algorithm was implemented in the rpart package [24], which is based on [25]. The target in the input dataset for the CART algorithm is numeric class, and the features are the n-dimensional vectors that represent attribute values of the target. In this work, cases per day of the group were taken as the target class, the features of which are shown in Fig. 2. CART creates a binary decision tree by the use of recursive binary splitting. As a result, the regression binary tree is built; the internal nodes of the tree represent the chosen attributes, and the left nodes show the number of cases per day.

The RandomForest package in R was employed to implement the random forest algorithm [26], which is based on [27]. The input dataset was the same as for the CART algorithm. In this study, we used the random forest to rank the importance of features [28]. As a result, the importance of each feature was calculated.

E. Outcome Definition and Statistics Methods

For the C5.0 algorithm, we defined a high risk of sICH as being more than the average number of cases daily from 2001 to 2011, the number of cases per day in the younger, middle-aged and older patient groups being 0.77, 2.26 and 2.60 cases, respectively (Table 1). The Student t-test was used for the continuous data, and the χ² test was used for the categorical data. ANOVA was used to assess the correlations between monthly mean weather factors and monthly mean sICH admissions. Statistical analysis was conducted using SPSS version 19.0 (SPSS Inc., Chicago, IL, USA). Statistical significance was defined as p < 0.05.

III. RESULTS

In total, 22,684 sICH cases were included in this study; 3,102 patients were younger (<44 years, younger group), 9,089 were middle-aged (45–64 years, middle group), and 10,457 were older (>65 years, older group). The proportion of low-income patients did not differ significantly between groups. As is usual, there was a significantly higher proportion of female patients in the older group than in the other age groups (p < 0.001) [1] (Table 1).

This study found that the number of cases of sICH per month was significantly highly negatively correlated with the average temperature ($R^2 = 0.851$, $p < 0.001$, correlation = −0.65), average maximum temperature ($R^2 = 0.877$, $p < 0.001$, correlation = −0.65) and average minimum temperature over the month ($R^2 = 0.812$, $p < 0.001$, correlation = −0.64). Regarding the average number of sICH cases per month, two other weather factors were examined, and a moderate positive correlation with the average wind speed ($R^2 = 0.394$, $p < 0.05$, correlation = 0.33) and a negative correlation with the average humidity during the month ($R^2 = 0.370$, $p < 0.05$, correlation = 0.14) were identified (Fig. 3).

Table 1. Demographic data of the sICH patients in Taipei City

<table>
<thead>
<tr>
<th>Age group</th>
<th>Younger (&lt;44 yr)</th>
<th>Middle (45–64 yr)</th>
<th>Older (&gt;65 yr)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cases</td>
<td>3,102</td>
<td>9,089</td>
<td>10,457</td>
<td>22,648</td>
</tr>
<tr>
<td>Number of cases per day*</td>
<td>0.77</td>
<td>2.26</td>
<td>2.60</td>
<td>5.63</td>
</tr>
<tr>
<td>Female (%)</td>
<td>29.88</td>
<td>32.78</td>
<td>43.49</td>
<td>37.3</td>
</tr>
<tr>
<td>Low income (%)</td>
<td>1.8</td>
<td>2.1</td>
<td>1.7</td>
<td>1.9</td>
</tr>
<tr>
<td>Total length of stay (day)</td>
<td>17.8 ± 15.3</td>
<td>18.2 ± 14.8</td>
<td>19.2 ± 15.8</td>
<td>18.3 ± 15.36</td>
</tr>
<tr>
<td>ICU length of stay (day)</td>
<td>5.4 ± 6.0</td>
<td>6.3 ± 6.8</td>
<td>6.6 ± 7.0</td>
<td>6.3 ± 6.8</td>
</tr>
<tr>
<td>Hospital mortality (%)</td>
<td>14.3</td>
<td>13.9</td>
<td>16.9</td>
<td>15.3</td>
</tr>
<tr>
<td>Total medical fees (US$)</td>
<td>6,026.1 ± 9,843.7</td>
<td>4,699.0 ± 5,603.3</td>
<td>4,907.0 ± 5,896.6</td>
<td>4,977.0 ± 6,361.1</td>
</tr>
<tr>
<td>Doctor fees (US$)</td>
<td>270.5 ± 250.8</td>
<td>254.8 ± 204.6</td>
<td>271.6 ± 225.4</td>
<td>265.0 ± 220.1</td>
</tr>
<tr>
<td>Room fees (US$)</td>
<td>1,610.6 ± 1,928.7</td>
<td>1,470.3 ± 1,590.5</td>
<td>1,599.2 ± 1,835.2</td>
<td>1,551.0 ± 1,756.9</td>
</tr>
<tr>
<td>Examination fees (US$)</td>
<td>338.5 ± 529.8</td>
<td>256.1 ± 367.1</td>
<td>309.0 ± 386.3</td>
<td>296.0 ± 401.1</td>
</tr>
<tr>
<td>Surgical fees (US$)</td>
<td>1,671.9 ± 1,324.7</td>
<td>1,335.4 ± 1,007.2</td>
<td>1,142.3 ± 864.5</td>
<td>1,317.0 ± 1,033.0</td>
</tr>
<tr>
<td>Medication fees (US$)</td>
<td>1,040.9 ± 6,528.7</td>
<td>767.4 ± 1,818.3</td>
<td>820.8 ± 2,230.0</td>
<td>829.0 ± 2,982.5</td>
</tr>
</tbody>
</table>

Values are presented as mean ± standard deviation.

*Calculated from 2001 to 2011.
This study evaluated the significant decision points of weather factors using C5.0, CART and the random forest method. Figs. 4 and 5 show the results of the C5.0 algorithm and CART algorithm, respectively, and Table 2 presents the results of the random forest method.

There was no significant influence of temperature, wind speed or humidity on sICH in the younger group (Figs. 4(a) and 5(a)). Two decision points were identified for patients in the middle group by C5.0: there was a higher risk of sICH if the maximum temperature (max_temp) was <19°C and the wind speed was <2.788 (m/s) (probability = 60.9%) (Fig. 4(b)). Otherwise, only one decision point was identified for patients in the middle group by CART: there was a higher risk of sICH if the maximum temperature in the second slot (max_temp2) was <23°C. Only one decision node was identified in this study for the older group: there was a higher risk of sICH if the average temperature (ave_temp) was <24°C (probability = 55.08%) (Figs. 4(c) and 5(c)). The top 3 most important factors in relation to sICH patients of different ages in Taipei City are shown in Table 2. The average wind speed and the average temperature are important factors affecting sICH.

IV. DISCUSSION

This study used a modified study design with different information technologies and re-defined weather factors in order to measure correlations with the incidence of sICH sequentially. Researchers have previously evaluated the effect of temperature using the average daily temperature [3, 5]. For example, temperature factors have been found...
Use of ITs to Explore Correlations between Climatic Factors and sICH in Age Group

Hsien-Wei Ting et al. 147 http://jcse.kiise.org

This study found that the maximum temperature and wind speed are risk factors that influence the incidence of sICH. Although the results obtained using decision trees differed slightly, it was concluded that similar results were obtained when employing the different information technologies.

The INTERACT2 study defined an ambient temperature <10°C as “cold weather”, and cold weather increases the risk of sICH [30]. However, older people have a higher risk of sICH in “warmer” cold weather. In our decision tree model, the older group had a higher risk of sICH if the mean temperature in the first 24 hours was lower than 24°C (real data = 23.933°C). Elderly people have a poor ability to defend against changes in weather, and even relatively warmer weather will place this population at higher risk of sICH. The elderly therefore need to maintain a warm environment to lower the risk of sICH.

This study found that different age groups are affected by different weather risk factors. Middle-aged patients have their own different risk models in terms of weather factors. The apparent temperature is correlated with humidity and wind speed [31], and researchers have previously found that the apparent temperature influences a person’s health if they work outside frequently, especially in high-heat environments [31, 32]. This study found that the apparent temperature has an influence not only in hot environments, but also in cold environments. Our results also showed that the apparent temperature influences the risk of sICH in middle-aged patients. A possible reason for this is that middle-aged people will often be the breadwinners in the home, and need to go outside to work; hence, the wind speed/apparent temperature significantly influence the risk of sICH in middle-aged people. In contrast, older people stay at home for more of the time, and so the influence of the apparent temperature is more important in middle-aged populations than in other age groups in relation to the risk of sICH.

The most common cause of young adult sICH is hypertension [33]; however, there are also other diseases or reasons that can cause sICH. For example, young adult

Fig. 5. Decision tree regression (CART) of sICH patients. (a) Aged below 45 years, (b) aged between 45 and 64 years, and (c) aged over 65 years. The internal nodes of the tree represent the chosen attributes, and the numbers at the top of the leaf nodes are the average number of sICH patients per day. N in the nodes gives the number of days.

Fig. 6. Decision tree regression (CART) of sICH patients aged below 45 years.
patients with arteriovenous malformation (AVM) or coagulopathy can develop sICH [33, 34]. This study found that weather factors do not influence the risk of sICH in young adult patients, no matter the temperature, humidity or wind speed. This study also identified a seasonal correlation with the number of sICH cases. The mean temperature was significantly negatively correlated with the number of sICH cases per month. The mean humidity and wind speed were also correlated with the number of sICH cases per month. The lower the ambient temperature and apparent temperature, the greater the risk of sICH, especially in the elderly. The weather is not the most important factor to consider in the prevention of sICH in young adults, but is very important in middle-aged and elderly populations.

This study used different information technologies, and although slightly different results were obtained, on the whole, they represented the same outcomes. However, this study had some limitations. First, the weather recording stations were not necessarily located in the area

Fig. 7. Decision tree regression (CART) of sICH patients aged between 45 and 64 years.

Fig. 8. Decision tree regression (CART) of sICH patients aged over 65 years.
in which the patients suffered sICH, and not all patients were sent to hospital at once. However, as most sICH patients are found and sent to the closest emergency room as soon as possible, the results of this study may still be considered to describe the relationships between weather factors and the incidence of sICH. Second, this study did not have access to data regarding the severity of sICH or the ICH area; therefore, we were unable to compare the severity under different weather conditions. Third, this study focused on Taipei City, and although the weather varies in the city, it is not a large enough area for analysis. Actually, these results were discovered from Taiwan’s database. It is reasonable that “warm” cold weather influences the risk of sICH patients in other countries. In the future, data could be collected using wearable equipment for greater accuracy, and weather data in different areas and countries could be obtained. Further study is required to confirm the reasons for which ambient temperature and “warm” cold weather influence the risk of sICH. We plan to conduct a large cohort study involving other datasets in the future, such as open government data and other patient data, and will employ big data analysis to perform a complete analysis of the risk of sICH.

V. CONCLUSION

First, this study found that weather factors have different influences on the risk of sICH in different age groups. There were no significant correlations of the risk of sICH with weather factors in the younger group. In the middle-aged group, the risk of sICH was found to be highly-correlated with the apparent temperature, and in the older group, the mean ambient temperature was found to be correlated with the risk of sICH. Second, although cold weather results in a higher risk of stroke, the results of this study showed that a “warm” cold ambient temperature still leads to a higher risk of sICH, especially in the elderly. Third, the decision tree method was found to be a good information technology for use in assessing the risk factors associated with medical issues. The method is simple to perform and the results are easy to understand.

ACKNOWLEDGMENTS

This study was partially supported by the Ministry of Science and Technology, MOST 104-2218-E-155-004 (funding received by CLC); MOST-105-2221-E-155-041-MY3 (funding received by CLC) and MOST 106-2218-E-155-007 (funding received by RKL). This study was based in part on data from the National Health Insurance Research Database provided by the National Health Insurance Administration, Ministry of Health and Welfare and managed by the National Health Research Institutes. The interpretation and conclusions contained herein do not represent those of the National Health Insurance Administration, Ministry of Health and Welfare or the National Health Research Institutes.

REFERENCES


Use of ITs to Explore Correlations between Climatic Factors and sICH in Age Group

Hsien-Wei Ting

Hsien-Wei Ting is now the Neurosurgeon and also the Chief of the Department of Evidence Based Medicine in Taipei Hospital, Ministry of Health and Welfare, Taiwan. He also is Adjunct Assistant Professor of the Computer Science and Engineering, Yuan Ze University. He also is both the Executive Secretary of Institutional Review Board and Medical Ethics Committee. He holds the MBA and Ph.D. of information management with the major of medical decision-making and medical informatics. He has published at least 20 journal articles for these majors.

Chien-Lung Chan

Chien-Lung Chan is now the Professor in Department of Information Management, and Innovation Center for Big Data and Digital Convergence, Yuan Ze University. He has published more than 30 journal articles in the field of Healthcare and Medical Informatics, Information Management, e-learning and Healthcare Services Research in recent 5 years. In 2013, he got the research excellence award from Yuan Ze University. He got his Ph.D. in Department of Industrial Engineering, University of Wisconsin-Madison, 1995.

Ren-Hao Pan

Ren-Hao Pan received Ph.D. degree in computer science from the National Chung Hsing University, Taiwan in 2014. He joined the department of computer science and engineering, Yuan Ze University, Chung-Li, Taiwan. Dr. Pan also works for National Yang-Ming University, Taipei, Taiwan. His research interests are in data engineering and big data analytics in medical informatics & bioinformatics.

Robert K. Lai

Robert K. Lai is a Professor of Computer Science and Engineering, Yuan Ze University, Taiwan. His current research interests include big data analytics, internet of things, and cloud computing. He received the Ph.D. degree in Computer Science from North Carolina State University in 1992.

Ting-Ying Chien

Ting-Ying Chien received his Ph.D. degree in Computer Science and Information Engineering from National Taiwan University in 2013. Now, he is an assistant professor at Computer Science and Engineering, Yuan Ze University, Taiwan. His primary research interests primarily cover big data analytics, bioinformatics, medical informatics, and big data systems system architecture.