

気象パターンの分類に基づく  
脳梗塞の発症と気象要素の関係について  
(Relationship between meteorological elements and onset of  
cerebral infarction based on the classification of weather pattern)

小島 光博  
(KOJIMA, Mitsuhiro)

名古屋大学大学院環境学研究科  
博士 (理学)

2014 年

# 要旨

季節や気象が病気を発症させる引き金となっている可能性について古くから研究されている。脳卒中や心筋梗塞などは、気象要素や気象変化によって発症する可能性が示唆されている疾患であり、気象病と呼ばれている。また、特定の季節に多く発症したり、死亡したりする病気は季節病と呼ばれている。このような季節や気象が病気の発症に及ぼす影響を研究する学問を生気象学という。脳卒中は代表的な季節病・気象病である。

ドイツでは古くから気象要素と脳卒中の関係についての研究が行われ、ハンブルクの医学気象予報 (bioweather) は有名である。ドイツでは今でも医学気象予報を配信している。近年では、世界各地で気象要素が脳卒中の発症に及ぼす影響を、入院患者や救急搬送記録を用いた疫学調査や、動物実験により研究されている。日本では松村氏などによって救急搬送記録を用いた疫学研究が行われている。その結果、脳卒中の発症は冬季に多く、気温が上昇するときや、気温が高いときにも発症が多いことが報告されている。しかし、これまでに気圧配置や前線の場所との関係に関する詳細な報告はない。

本研究では脳卒中の一種である脳梗塞に注目し、名古屋市冬季における脳梗塞の発症に及ぼす気象要素の影響について、天気図型 (気候パターン) を考慮した新しい統計的手法を提案する。また、それぞれの気候パターンにおいて、脳梗塞の発症に影響する気象要素の特定と、発症の予測可能性について検討を行うため、予測精度も算出した。

本研究では、名古屋市消防局の救急搬送記録から脳梗塞として搬送された患者を抽出し、気象庁から名古屋気象台の気象情報を取得してデータベースを構築した。気象要素は多重共線性の問題を回避するため主成分分析を行い、取捨選択を行った。

脳梗塞の発症に影響する気象要素を定量的に分類し、気候パターンの分類を行うために、Self-Organize Mapping (SOM) を用いた。SOM は、ニューラルネットワークの一種であり、教師なしの競合学習により多次元データを二次元に分類・可視化することに効果的であるため、工学・医学・気象学などの分野に応用されている。本研究では、SOM を発展させた、教師ベクトルを持つ双方向 SOM を用いた。

また、気象要素と脳梗塞の発症の関係について統計的な解釈をするために、サポートベクターマシン (SVM) と決定木を用いて研究を行った。SVM はマージン最大化と非線形モデルの特徴を活かして、汎用性と分類能が高い統計手法として知られている。決定木はデータの属性について分割する基準を与える続けることで結論をつけることができるため、可読性の高い分類手法として多く用いられている。

これまでの先行研究では、それぞれの気象要素と脳梗塞の発症数の関係について統計解析していたため、複雑な気象変化には十分に対応していなかった。

本研究では双方向 SOM を用いることにより、脳梗塞の発症に関係する天気図として、「雨天」「晴天」「西高東低」の3つの気候パターンの分類に成功した。さらに、SVM と決定木により気候パターンごとに脳梗塞の発症に影響を与える気象要素の特定に成功した。また、気候パターンを分類したことで、分類しなかった場合よりも 10% 近く精度が向上していることが確認された。また、非線形モデルの SVM を用いたことで、線形モデルの決定木では解釈が難しかった脳

梗塞の発症に与える気象要素の影響を視覚的にも容易に理解できるようになった。

本研究により気象情報を基にした脳梗塞の発症予測モデルの構築に成功し、脳梗塞の発症予測について有効な発症予測モデルの開発に成功した。

## Abstract

There have been many studies that discussed the possibility that a season and the weather trigger the onset of many kinds of diseases. A stroke, myocardial infarction and so on, are called “meteorological disease”, because possibility of their onset by meteorological change was suggested. In addition, disease that appears in a specific seasons, is called “climatic disease”.

Biometeorology is a field of study that investigates the effects by a season and the climatic elements against disease. The stroke is regarded as a typical climatic disease and meteorological disease.

In Germany, there are many studies of relationship between climatic elements and onset of stroke for a long time, and they developed well-known “Bio-weather forecast (health-weather forecasting systems)”. “Bio-weather forecast” is still announced in Germany. In these years the influence that meteorological elements effect s to stroke onset is studied in all parts of the world by an epidemiology and the animal experiment using databases of inpatient and the medical emergency transfers.

The epidemiologic study using the database of medical emergency transfers was studied by Dr. Matsumura and others in Japan. Then, there are reports that the onset of the stroke has, peaks during the winter and when temperature is high or rises. But, the detailed report about relationship between the onset of stroke and distribution of atmospheric pressure or place of the front has not yet been done so far.

In this study, we focus on cerebral infarction (CI) that is a kind of stroke, and propose new statistical methods considering weather map type (climate pattern) about the effects of climatic elements on the onset of CI in Nagoya during the winter. In addition, we calculated the predictive precision to examine the identification of meteorological elements to influence the onset of the CI in each climate pattern and predictive possibility of the onset.

In this study, we select transformed patients that had been diagnosed CI, from the records of medical emergency transfers of Nagoya City Fire Department, then we took the climatic information of the Nagoya Meteorological Observatory from Japan Meteorological Agency and constructed databases. We avoided the multicollinearity problem by performing a principal component analysis of the meteorological elements. We applied a Self-Organizing Map (SOM), to classify quantitatively the effects of climatic elements on the onset of CI.

The SOM has been applied in various fields including meteorology, medicine and engineering. It is a kind of neural network and effective in classifying / visualizing complex high-dimensional data on simple two-dimensional grids by unsupervised competitive learning.

In this study, we applied bi-directional SOM (with a teacher vector) that had been developed from the SOM.

On the other hand, in order to interpret statistically relationship between onset of CI and meteorological elements, we investigated using the method of Support Vector Machine (SVM) and decision tree. Decision tree is often applied as high-legibility classification method.

About relationship between onset of stroke and climatic elements (particularly atmospheric pressure and temperature), Hiroshima prefectural medical association predicted risk of onset by Poisson Regression Analysis using the records of medical emergency transfers

We classified successfully a daily weather maps into three climatic patterns, such as “rainy weather”, “fair weather” and “west high and east low pressure” by applying bi-directional SOM. Furthermore, the identification of meteorological elements which influenced each climate pattern was performed SVM and decision tree.

It was confirmed that nearly 10% of precision by SVM of the non-linear model, compared with the decision tree of linear model easily visually. In addition, we came to understand the mutual relations of the meteorological elements that interpretation was difficult with the decision tree of linear model easily visually by having used SVM of the non-linear model.

This paper proposed a new method by SVM to predict the onset of cerebral infarction with high accuracy.

# 目次

<b>第1章 序論</b>	<b>1</b>
1.1 背景	1
1.2 生気象学と医学気象予報	2
1.3 海外の医学気象予報事例	4
1.4 本研究の内容	5
<b>第2章 脳梗塞発症データベースおよび、 気象情報データベースの構築</b>	<b>9</b>
2.1 脳梗塞発症データベース	9
2.1.1 救急搬送について	9
2.1.2 消防局の救急搬送データの概要	10
2.2 気象情報データベース	12
2.2.1 気象庁気象情報データベースについて	12
2.2.2 解析用気象情報データベース	13



<b>第3章 解析データベースの構築及び解析手法</b>	<b>16</b>
3.1 気象情報データベースの気象要素間における多重共線性の回避	16
3.2 双方向 SOM による気候パターンの分類	17
3.3 決定木とサポートベクターマシン (SVM) による 脳梗塞の発症予測モデル構築と予測精度の検証	19
<b>第4章 気象パターンの分類と、 脳梗塞の発症リスクの結果と考察</b>	<b>21</b>
4.1 はじめに	21
4.2 方法	22
4.3 結果と考察	24
4.3.1 気象パターンの分類	24
4.3.2 脳梗塞の発症リスク	27
<b>第5章 発症リスクの予測精度</b>	<b>30</b>
5.1 はじめに	30
5.2 方法	30
5.3 結果	32
5.3.1 研究対象期間の発症予測精度	32
5.3.2 気象パターンごとの発症予測精度	33

<b>第 6 章 結論と課題</b>	<b>37</b>
6.1 結論	37
6.2 展望と課題	40
謝辞	42
引用文献	43

## Appendix

<b>第 7 章 解析手法</b>	<b>50</b>
7.1 SOM (自己組織化マップ : Self-Organizing Map)	50
7.1.1 SOM	50
7.1.2 双方向 SOM	52
7.2 決定木分析	54
7.3 SVM (サポートベクターマシン : Support Vector Machine)	57
図表	59



## 第1章 序論

### 1.1 背景

日本には、「低気圧が近づくと頭痛がする」、「季節の変わり目には体調を崩す」、「雨が降る前からぜんそくが悪化する」、「雨が降ると古傷が痛む」など、気象が人体に及ぼす影響について慣習的な認識がある。このような影響が広く認められ、西洋医学の発展に伴った経験や科学的な根拠に基づく医療が浸透してきたことで、医療現場では患者に関する詳細なデータが収集されてきた。これらのデータに基づいて、季節や気象が病気を発症させる引き金となっている可能性について、統計的な解析をもとに論じられるようになった（加地ら, 1975）。

脳卒中・脳梗塞・心筋梗塞などの循環器系疾患や、ぜんそくなどの呼吸器系疾患が、気象要素（気温・気圧・湿度）やそれらの変化（前日との気温差、雨天や晴天）が発症の引き金となっている可能性が示唆されている。これらの疾患は「気象病」と呼ばれている。また、夏季に多発する熱中症など、特定の季節に集中して発症し、時には死亡に結びつく疾患は「季節病」と呼ばれている。季節や気象が病気の発症に及ぼす影響を研究する学問分野は、「生気象学」と言われている。

ドイツではヒポクラテスの時代（紀元前5世紀ごろ）に既に気象要素と脳卒中の関係についての研究が行われ、ハンブルクで医学気象予報（bio-weather）が開始された。ドイツでは現在でも、民間気象会社がホームページやテレビで医学気象予報を配信している（Goggins *et al.*, 2012）。

近年では、世界各地で気象要素・気象変化が脳卒中の発症に及ぼす影響を探るため、病院の入院患者記録や救急搬送データを用いた疫学調査や、気圧や気温、湿度などを人工的にコントロールできる特殊な施設を用いた動物実験など

が多くの研究機関で行われている (Sato *et al.*, 1999).

これまでの研究から、脳卒中の発症は年間を通じて気温が最も低くなる冬季に多発する傾向があることが示唆されている (Walach *et al.*, 2002; Inagawa *et al.*, 2002; Fisher *et al.*, 2004). また、気温の上昇時や、気温が高いときにも病気の発症が多いことが報告されており、気温が脳卒中の発症に及ぼすことがほぼ確実である (例えば, Chen *et al.*, 2013; Kawai *et al.*, 2001; Wang *et al.*, 2006; Lim *et al.*, 2013).

これまでの研究では、複雑に相互連携している気象要素と病気の発症の関係について総合的な解析が十分とは言い難い。また、気象要素が病気の発症に与える影響についての研究は、特定の気象要素と病気に発生数 (患者数) との相関関係や有意差検定などで調べられているだけで、直観的あるいは経験的な関係しか論じられていないのが現状である (Setzer *et al.*, 2007; Hong *et al.*, 2003; Woo *et al.*, 1991). 脳梗塞の発症に与える気象要素の影響についてより厳密な解析をするには、相互に複雑に関係している気象要素を総合的に検討し、気象概況に直接的な影響を与える気圧配置や前線の場所との関係まで包括的に評価する必要が論じられている。相互に複雑に関連している気象要素と病気の発症の直接、あるいは間接的な関係を考えることは、今後の重要な課題とされている。

## 1.2 生気象学と医学気象予報

国際生気象学会 (International Society of Biometeorology) によると、「生気象学」は、大気の物理的、化学的な条件 (気象要素や気象変化・気象条件、季節変化) が人間・動物・植物に及ぼす直接、間接的な影響を研究する学問と定義されている。

気象要素や気象変化によって発症する可能性が示唆されている疾患は「気象病 (meteor tropical disease)」と呼ばれる。気象病にはリウマチ、脳卒中、脳梗塞、ぜんそく、心筋梗塞などが挙げられる。また、特定の季節に多く発症したり、死亡したりする疾患は「季節病 (seasonal disease)」と呼ばれている。冬の季節病には脳卒中、心臓疾患などの循環器系疾患や、気管支炎などの呼吸器系疾患があり、夏の主な季節病には熱中症・脱水症が挙げられる (Fukuoka *et al.*, 2003)。また、近年では家庭内の転倒・転落事故や溺水事故などの家庭内事故についても気象条件が2次的に影響を及ぼしているという報告もされている (Uesaka *et al.*, 2003)。

これらの季節病・気象病・家庭内事故の発症数 (発生数) と気象情報の関係について統計学的な解析を進めることで、これらの発症 (発生) する危険度 (リスク) に応じた注意報や警報として予報配信するシステムを構築することができ、予防・抑制・低減に繋がると期待されている。このようなシステムは医学気象予報や健康天気予報と言われる。WHO (世界保健機関) によると、気候変動により熱中症・脱水症などの暑熱関連疾患、マラリアなどの熱帯感染症および精神系疾患などが増加すると予測している。これらの対策として、Health Weather Watch/Warning System (健康・気象・監視/警報システム) の構築・普及の重要性が提唱されている (Pascal *et al.*, 2006)。

### 1.3 海外の医学気象予報事例

ドイツでは、約70年前から気象病と季節病とされる約30疾患（脳卒中、心筋梗塞、ぜん息、リウマチなど）の発症リスク（疫学的・統計学的に特定の疾患について発症の恐れがある危険度）を4段階評価で表現した「医学気象予報 (bio-wetter)」が、毎日配信されている (Goggins *et al.*, 2012).

1969年当時の医学気象予報、「天気の状態と天気予報：雷雨前線の後方にあつて、冷たい海洋気団の影響を受け、当地方は低気圧の変わり易い天気である。南バイエルンは風下の効果により不安定の度が強くなっている。医師へ：自覚症状は痛感がやや増加する。病気への影響は昼間だけ痙攣が増すがその他による影響はない」は、その一例である。一般人に向けて医学気象予報を配信し、注意喚起をすることで病気の発症予防と低減を目的にしたものである。また、医師に対しての報告もあり、臨床の現場で参考とされている。時間帯や特定の疾患についてのコメントが付記され、勤務体制や患者の受け入れ準備において参考となる情報が配信されている。

WHOが提唱している Health Weather Watch/Warning System は、すでにフィアデルフィア、ローマ、リスボン、上海などの大都市において構築され、運用が開始されている。フィアデルフィアでは熱波（気温35°C以上が数日間継続する状況）の接近が予報されると、熱波警報をメディアにより発令し、市民への広報活動や自治体の保健局などの職員やボランティアによる電話相談、高齢者向けの訪問指導、電気・水道などのインフラの確保、冷房完備施設の設置など多種多様な対策がとられている。フランスでは2003年8月の熱波が到来したことで、高齢者を中心に1万5000人が脱水症などで死亡した。これによりフランス政府やパリ市は熱波対策が遅れたことについて、市民から批判された (Pascal *et al.*,

2006; Kovats *et al.*, 2006). 最近, 日本においても医学気象予報に対する注目は高まっており, テレビ番組内の天気予報において熱中症, UV・紫外線情報, 花粉飛散状況が報道されるようによくようになった.

近年, 地球温暖化や気候変動の影響により熱帯地域に多く生息していた媒介虫によるマラリアなどの感染症が世界的に拡大する可能性についても議論が始まっており, 医学気象予報の重要性が世界的に再認識されるようになった (Martens *et al.*, 1999).

#### 1.4 本研究の内容

厚生労働省が発表している「人口動態統計の概況」によると, 平成 25 年における年間の死因別死亡総数のうち脳血管疾患は 11 万 8347 人であり, 死亡数で上位 4 位 (全死因の 9.3%) を占めている. このうち脳梗塞は 6 万 9967 人あり脳血管疾患の約 6 割が脳梗塞で死亡している. 脳梗塞は発生数がとても高いことから, 本研究の対象疾患とした.

本研究は, 脳梗塞の発症を気象要素から高精度に予測できる「発症予測モデル」を構築し, 脳梗塞の発症予測に有効な手法の開発を目的としている.

これまでの先行研究では, 気温などのそれぞれの気象要素が脳梗塞の発症に及ぼすリスクについて比較検定や相関分析により解析されてきた (Walach *et al.*, 2002; Inagawa *et al.*, 2002; Fisher *et al.*, 2004; Chen *et al.*, 2013; Kawai *et al.*, 2001; Wang *et al.*, 2006; Lim *et al.*, 2013).

しかし, 病気の発症には気温などの特定の気象要素が影響しているのではなく, 複数の気象要素が複雑に影響していると予察され, 単純な統計解析では脳梗塞の発症と気象要素の関係について十分な知見が得られているとは言い難い.



本研究では、気象要素や気象パターンをタイプ分けすることで、脳梗塞の発症についてより総合的な統計学的な解析法を提案し、それを応用し名古屋市における脳梗塞の発症リスクについてより厳密な検討を行った。また、脳梗塞の発症リスクについての予測可能性について論じた。

まず、研究対象期間である2002-2004年の冬季（12月から翌年2月）における名古屋市の気象要素について検討を行った。気象データは気象庁名古屋気象台が発表している気象情報（平均気圧、平均気温、最低気温、最高気温、平均相対湿度、平均蒸気圧、平均雲量、降水量）を用いた。また、これまでの研究から脳梗塞の発症には前日の気象要素の影響がするとの報告(Setzer *et al.*, 2007 ; Hong *et al.*, 2003 ; Woo *et al.*, 1991)がなされていることを考慮し、本研究では気温日較差（当時における最高気温と最低気温の差）や平均気圧前日差（当日の平均気圧と前日の平均気圧の差）、平均気温前日差（当日の平均気温-前日の平均気温）など、気象要素の変化についても検討を行った。

次に、これらの気象情報について主成分解析を行い、多重共線性の問題を回避した。このように脳梗塞の発症に与える気象要素の関係を広汎に見るため新たに導入した気象情報の指標を加え「気象情報データベース」を構築した。

本研究で用いる脳梗塞の発症数に関するデータは、名古屋市内を管轄している名古屋市消防局の救急搬送データを用いた。この救急搬送データには、搬送ごとに、日時、病名、場所、性別、年齢、傷病程度などが詳細に記載されている。この救急搬送データから、病名に「脳梗塞」として記録されているデータを抽出することで脳梗塞の発症件数を集計し、名古屋における脳梗塞の発症件数を時系列で参照する「脳梗塞データベース」を構築した。

「気象情報データベース」には多くの気象要素が多次元で存在している。また、「脳梗塞データベース」には脳梗塞の搬送数が時系列で存在している。

そこで、多次元のデータベースを二次元平面上に可視化できるニューラルネットの1つである双方向自己組織化マップ(双方向 SOM: Bi-directional Self-Organizing Map) を用いた. 双方向 SOM は教師ベクトルと入力ベクトルがあり, 競合層のユニット ( $n \times m$ ) に対して双方向から, 競合層のそれぞれのユニットに対して重みベクトルが学習されて二次元平面上に学習される. このように, 双方向 SOM を用いることで競合層のユニット上に双方向の重みベクトルが視覚的に表現でき, 理解が促される. このユニットに学習されている重みベクトルを解析することで, 多次元のデータベースについて特徴を把握することができる.

本研究では, 教師ベクトルを「気象情報データベース」とし, 入力ベクトルを「脳梗塞データベース」にして双方向 SOM を求め, 二次元平面である競合層のユニットとして学習させた. つまり, 競合層には「気象情報データベース」にある多次元の気象要素について重みベクトルが学習され, 「脳梗塞データベース」の脳梗塞の搬送数も同じ競合層に学習されているために脳梗塞の発症リスクと気象要素との関係を競合層のユニット上に可視化できる.

次に, 競合層に学習されている「気象情報データベース」のユニットについて, 階層型クラスター分析を行った. これにより気象学的に近いグループを分類することができ, 「気象パターン」を複数に分類した. 分類された気象パターンをグループ化した(本研究では, グループ A, グループ B, グループ C とする). これにより気象学的に類似したケースが気象パターンとして分類される.

これらの気象パターンが, それぞれどのような特徴や傾向を有しているか2つの方法で確認した. まず, 気温 ( $^{\circ}\text{C}$ ) や降水量 (mm) などの気象要素とそれぞれの「気象パターン」を比較することで, 気象学的に類似したケースの特徴や傾向が説明できるか検証した. さらに, 一般財団法人日本気象協会によって,

前線の位置や気圧配置などの気象学的な基準で作成した指標で分類された「天気図分類表」と「気象パターン」との比較を行った。つまり、異なる方法により、双方向 SOM で分類された「気象パターン」の特徴や傾向が正しく表現できているかについての検証を行った。

以上のような手法を用いて「気象パターン」に分類することができるため、それぞれの「気象パターン」における脳梗塞の発症リスクについての研究を行った。本研究では脳梗塞の発生頻度が多い日から「リスク・大」「リスク・中」「リスク・小」とし、対象日数を均一になるように設定した。脳梗塞の発症リスクをより明確に比較するために「リスク・大」と「リスク・小」のみを抽出した。

ここでは脳梗塞の発症リスクに与える気象要素の特徴を確認するだけでなく、気象要素による脳梗塞の発症予測についても検討を行った。つまり、「気象パターン」ごとに従属変数を気象要素とし、目的変数を脳梗塞の発症リスク（「リスク・大」と「リスク・小」として、決定木学習アルゴリズムを使った分析とサポートベクターマシン（SVM）による分析を行った。決定木学習アルゴリズムを使った解析は、線形モデルの特徴から直線的な基準により脳梗塞の発症リスクを分類できる。SVM は非線形モデルとマージン最大化の特徴から「リスク・大」と「リスク・小」の境界を判別することが可能である。

決定木学習アルゴリズムを用いた分析と SVM により、それぞれの「気象パターン」において、発症リスクの「発症予測モデル」を構築することで、それぞれ精度評価を行った。また、双方向 SOM を用いない場合の発症予測モデルを構築し、「気象パターン」に分類した場合との比較を行った。

## 第2章 脳梗塞発症データベースおよび、気象情報データベースの構築

### 2.2 脳梗塞発症データベース

#### 2.1.1 救急搬送について

全国で、救急車は日平均で1万5000件が出動しており、平成24年度には過去最多の580万件（前年度比9.4万件増、1.7%増）の出動があった。これは5.8秒に1件の割合で救急車が出動していることになる。救急搬送は平成16年度に500万件を超えて以来、年々増加する傾向が認められる。最近10年間では約30%も増加している。このような背景には高齢者化や一人暮らし世帯の増加により、病状の急変や通院までに介護を要する利用者が増えていることが挙げられる（消防白書、2013年）。さらに、軽症患者による病院までのタクシー代わりに利用するなど不適切な利用者による出動が増えていることも指摘されている（消防白書、2013年）。一方で、救急搬送数が増え、救命救急医療の必要性が高まっている中、救急隊員（救命救急士）数は同期間で6%ほどしか増加していない。救急車が傷病者の受け入れ先の医療機関を確保するまでに長時間かかる場合や、救急車が不足している状況が想定される。最近では、たらい回しにより救急搬送中に死亡、重篤化することがあり、この状況の改善は、喫緊に解決しなければならない社会問題である。

将来の救急搬送件数の将来予測では、2010年以降は日本の総人口は減少に転じる一方で高齢化率が上昇することで、2030年までに救急搬送件数は増加傾向が続き600万件を超えるとされている（消防白書、2013年）。さらに、日本における医療費は増加を続けており、今後の社会保障費の見通しと医療費の増大が社会問題として、連日の報道でも大きく取り上げられている。これらを背景にして、消防局では予防救急という取り組みを始めている。

## 2.1.2 消防局の救急搬送データの概要

救急搬送データを用いた研究は、脳梗塞以外にもぜん息発作や急性心筋梗塞について報告されている (Hanashiro *et al.*, 1998; Wang *et al.*, 2007). 救急搬送データには、救急隊が救急車による救急搬送の依頼を受けて出動し、現場に到着して対象者を搬出し病院に搬送され、搬送先の救命救急医が初期診断するまでの活動の詳細が記録されており、その地域を所管している消防局が救急搬送データを保有している。つまり、「出動年月日」「出動時分」「学区」「小学校区」「傷病者年齢」「性別」「事故発生場所」「病名」「傷病名」「傷病程度」「負傷部位」が詳細に記録されている。本研究では、救急搬送データに含まれる氏名や住所など個人を特定できるような情報は、消防局の担当者により事前に除外されている救急搬送データを用いた。

救急搬送データは、国民や市民の健康維持促進や救命救急医療の発展、社会課題の解決、予防救急の施策に向けた研究に活用されている。例えば、独立行政法人国立環境研究所では政令指定都市の救急搬送データを活用し、日本における熱中症の実態およびリスク要因について研究を行っている (Ono *et al.*, 2009)。この研究では、都心部で熱中症の増加傾向がみられ、年々過去最多の搬送件数を記録していることが明らかとなった。また、日最高気温の上昇に伴って発症数が増加し、小中高生では最高気温が 35°C を超えてからは穏やかになる一方で、高齢者は上昇を続けていることから、熱中症の発症予防対策が十分に行われていないことを指摘された。

このように救急搬送データを解析は、熱中症予防に向けた指針や対策の策定に有効である。近年では、これらの指針に基づいた熱中症予防情報を環境省が注意報・警報の情報として配信し、国民に対してホームページなどを用いて注

意喚起している。また、救急搬送データには屋内の発生場所（浴室やトイレ）などの詳細なデータが含まれているため、日本建築学会では救急搬送データを用いた研究も実施されている。特に、救急搬送データの分析から浴室まわりにおける心不全の発症について多数の研究が行われ、ヒートショックにより心不全などが誘発される仮説が実証されつつある。これらの研究成果を基にして、浴室の温熱環境を計測機により実測することで、発症メカニズムや発症を抑制する提案を行っている。このように、救急搬送データを解析することで、病気の発生メカニズムの推定や、予防・抑制の提案が行える。

本研究では、名古屋市内においてはおよそ年間 10 万件的救急搬送があり、搬送ごとに詳細な救急搬送データが蓄積されている。2002 年から 2004 年に名古屋市消防局の管轄内で脳梗塞と診断されて病院に搬送された人を本研究の対象として、名古屋市消防局が保有する救急搬送データから抽出した。抽出されたデータから、発生日時を知ることができるので 1 日に何件が脳梗塞として搬送されたかを時系列で集計し、「脳梗塞データベース」を構築した。

2002-2004 年の名古屋市冬季における 1 日あたりの脳梗塞患者の救急搬送数のヒストグラムを図 1 に示した。2002-2004 年冬季の名古屋市において脳梗塞として搬送された患者数は 1435 名で、1 日当たり最低 1 名～最大 13 名であった。1 日当たりの平均は 5.3 名であり、5 名の患者が搬送された日が 55 日と最も多かった。

## 2.2 気象情報データベース

### 2.2.1 気象庁・気象情報データベースについて

気象庁は日本国内の的確な気象情報を提供することによって、自然災害の軽減、国民生活の向上、交通安全の確保、産業の発展などを実現することを任務としている。そのため、札幌管区気象台、仙台管区気象台、東京管区気象台、大阪管区気象台、福岡管区気象台、那覇に沖縄気象台を設置して気象観測などの業務をしている。これらの観測地点で広域的な気象、地震、海洋などの観測・監視、予報・警報や情報提供などを行っている。さらに地方気象台を設置することで、より詳細な情報を収集している。

環境省では熱中症などに対する注意を呼びかけるために、独自に熱中症予防情報サイト (<http://www.wbgt.env.go.jp/>) を運営している。このサイトでは、1954年からアメリカで採用されている暑さ指数（(湿球黒球温度) WBGT: Wet Bulb Globe Temperature) を用いて、日本各地に対して熱中症の警報や注意報を国民に広く注意喚起している。

WBGTは、気温(°C)だけに注目せず、人体における熱収支についても検討しており、気温以外にも湿度や日射なども予測式に含まれている。このように気象情報を用いることで、熱中症の予防に役立てることができることが明らかになっている。

気象庁が公開している気象要素は、平均海面気圧(hPa)、降水量(mm、合計降水量)、平均気温(°C)、最高気温(°C)、最低気温(°C)、平均湿度(%)、最小湿度(%), 平均風速(m/s)、日照時間(時間)、全天日射量(MJ/m<sup>2</sup>)、平均雲量(10段階)、降雪の深さ合計、(cm)、最深積雪(cm)、平均蒸気圧(hPa)の14要素である。

## 2.2.2 解析用気象情報データベース

本研究では 2002-2004 年の冬季（12 月から翌年 2 月）において名古屋市消防局が管轄している名古屋市で脳梗塞と診断された搬送件数と、この搬送件数に与える気象要素の影響について統計解析を行うことを目的としている。そこで、同期間に名古屋市内において気象観測をしている気象庁名古屋地方気象台（愛知県名古屋市千種区）が気象観測を行い、その結果を取りまとめた気象情報を基にした気象情報データベース (<http://www.data.jma.go.jp/obd/stats/etrn/index.php>) の構築を行った。気象庁はホームページで過去の気象情報を一般に公開しているため、研究対象期間における名古屋市の気象データを収集した。本研究では日単位 (00:00-23:59) で集計された気象情報の収集を行った。

これまでに報告されてきた大部分の論文では、当日の平均気温 (°C) や平均気圧 (hPa) などの気象要素が脳梗塞の発症に影響を及ぼしている可能性を示唆している (Setzer *et al.*, 2007 ; Hong *et al.*, 2003 ; Woo *et al.*, 1991)。「雨が降ると古傷が痛む」、「低気圧が近づいてくると頭痛がする」、「雨天が続くと気分が落ち込む」、「急激に温度が変化するとぜんそくの発作がおこる」など、刻々と変化する気象状態により体調や健康状態、気分的な落ち込みを感じるといった報告がある。また、先行研究から脳梗塞や脳卒中の発症には気圧の変化が有意に増加するという結果が報告されている (Ono *et al.*, 1974)。

本研究では、気象庁（名古屋気象台：愛知県名古屋市千種区）が発表している気象情報データから研究対象期間について、「対象日の平均海面気圧 (hPa)」と「対象日前日の平均海面気圧 (hPa)」を抽出し、その差分「平均海面気圧前日差 (hPa)」を気象パラメータとして気象情報データベースに追加した。対象日



前日より対象日の平均海面気圧(hPa)が上昇した場合は正值(プラス)になり、下降した場合には負値(マイナス)となる。

脳梗塞や脳卒中の発症に及ぼす気象要素の関係について報告されている研究の中で、もっとも多いのは気温に関連している。多くの先行研究では気温が低い場合に発症のリスクが高くなる傾向があるという報告や、気温が低い時と高い時の2つのパターンで死亡率が上昇する傾向が報告されている(Passero *et al.*, 2000; Shinkawa *et al.*, 1990; Feigin *et al.*, 2000; Wang *et al.*, 2003)。

本研究では、平均気温(°C)、最高気温(°C)、最低気温(°C)について、「対象日の気温(°C)」と「対象日前日の気温(°C)」を抽出し、それらの差分値をそれぞれ平均気温前日差(°C)、最高気温前日差(°C)、最低気温前日差(°C)とし、新しい気象パラメータとして気象情報データベースに追加した。対象日前日より対象日の気温(°C)が上昇した場合は正值(プラス)になり、下降した場合には負値(マイナス)となる。

また、日内の気温変化についても研究され脳梗塞の発症リスクに影響している可能性が示唆されている(Hasegawa *et al.*, 1985)。そのため、対象日の最高気温(°C)から同日の最低気温(°C)を差し引いて、気温日較差(°C)を算出し、気象パラメータとして気象情報データベースに追加した。気温日較差(°C)は日内の気温差が激しいと数値が大きくなり、気温差が少なければ数値は小さくなるため、日内変動も考慮しているため有益な指標性を持つと考えられる。

脳梗塞の発症には、当日の気象要素だけではなく、前日の気象変化や、前日と当日の急激な気象変化が脳梗塞の発症に影響を与える可能性を示唆している報告が多くある(例えば、Ono *et al.*, 1974)。これらの研究を踏まえ、本研究では、新たな気象パラメータとして、平均海面気圧前日差(hPa)、平均気温前日差(°C)、最高気温前日差(°C)、最低気温前日差(°C)、気温日較差(°C)を考慮した。

さらに、気象要素と脳梗塞の関係を研究するため、平均海面気圧前日差(hPa)と同様な算出方法により、平均湿度前日差(%), 平均風速前日差(m/s), 平均雲量前日差(10段階), 日照時間前日差(時間), 平均蒸気圧前日差(hPa)を気象情報データベースに追加した。これらの計算式は、表1に示す。

各気象要素の前日差は、気象庁(名古屋地方気象台)が発表している気象情報から、各気象要素において「対象日の実測値」と「対象日前日の実測値」を抽出し、それらの差分値である。対象の前日より実測値が上昇した場合は正值(プラス)になり、下降した場合には負値(マイナス)となる。

このように、これまでの研究で報告されている脳梗塞の発症に影響を与えると考えられる気象パラメータを使い解析を行った。脳梗塞の発症に影響を与える可能性がある気象要素が取り込まれた「気象情報データベース」を新たに構築した。

### 第3章 解析データベースの構築及び解析手法

#### 3.1 気象情報データベースの気象要素間における多重共線性の回避

一般的に統計学的手法を用いて予測を行う際、説明変数の種類を増やすことで予測精度は向上する傾向がある。また、統計解析に用いる説明変数の中に変数間の相関関係が強い変数が含まれている場合や、一次従属な変数関係がある場合に予測精度が低下することがある。この現象は多重共線性と呼ばれている。

多くの説明変数が存在している場合は、数的に変数間を区別することが困難で、変数を用いた因果関係や回帰式の解析を行う場合には多重共線性については考慮しなければならない。つまり、再現性や汎用性が低下することにより予測モデルの信頼が低下する可能性がある。

本研究では多重共線性を回避するために、主成分分析を用いて説明変数の取捨選択を行った（図2）。図2のように強い相関を示している、または一次従属の変数関係にある気象要素を主成分分析により判別して、気象要素を取捨選択した。これまでに構築した「気象情報データベース」について主成分分析を行った。累積寄与率が68.5%となる第二成分までを使用して多重共線性について取捨選択を行った。

平均相対湿度と平均雲量に強い相関が認められたため、平均雲量を破棄して平均相対湿度を採用した。また、全天日射量前日差と日照時間前日差にも強い相関が認められたため、日照時間前日差を採用した。このような多重共線性の問題を回避する操作を繰り返して、本研究で用いる気象要素を選択した。

本研究では、脳梗塞の発症に与える気象要素の影響を正確に把握するための検討を行い、気象要素を広くに把握することができる「解析用気象情報データベース」を構築した。多重共線性の検討後に本研究で用いた気象要素は、平均海

面気圧，平均気温，最高気温，気温日較差，平均相対湿度，日照時間，全天日射量，降水量，海面平均気圧前日差，最高気温前日差，平均相対湿度前日差，日照時間前日差であり，本研究ではこの 12 気象要素を脳梗塞の発症に影響を与える要素とした。

### 3.2 双方向 SOM による気象パターンの分類

季節や気象要素が脳梗塞の発症に及ぼす影響についての研究は昔から注目されている。これまでの研究は気温 (°C)・気圧 (hPa)・湿度 (%) などの気象要素が脳梗塞の発症に与える影響について，相関関係や回帰分析，比較検定など比較的単純な統計分析が行われていた (Setzer *et al.*, 2007; Hong *et al.*, 2003; Woo *et al.*, 1991). その成果として，冬季に脳卒中の発症が多く，気温が上昇するときや，気温が高いときにも発症が多い傾向が明らかとなった (Walach *et al.*, 2002; Inagawa *et al.*, 2002; Fisher *et al.*, 2004). しかし，気象要素が脳梗塞の発症に与える影響を総合的に判断して発症リスクの判別をしているものは未だ存在しない。また，気圧配置や前線の位置と脳梗塞の関連性に関する詳細な報告はない。

本研究は脳梗塞の発症リスクに影響する気象要素の特徴を総合的に把握するだけでなく，脳梗塞の発症リスクに関係する天気図型についての検討も行い，影響を及ぼしている気象パターンの特徴を把握することを目的とする。また，気象パターンに分類した結果についても正確な分類ができているか検証を行う。さらに，気象パターンと実況天気図を比較するだけでなく，日本気象協会東海支社が分類した天気図型分類表とも比較を行う。

本研究では、脳梗塞の発症に与える気象要素の影響を総合的に把握するために、多次元のデータベースを二次元平面上に可視化できる双方向 SOM を用いた。双方向 SOM は教師ベクトルと入力ベクトルがあり、競合層のユニット ( $n \times m$ ) に対して重みベクトルが学習され、それぞれのユニットに重みベクトルが学習されている。つまり、二次元平面上に競合層のユニットを学習することができる。

本研究では、競合層には  $5 \times 5$  の 25 個のユニットを設定した。「気象情報データベース」の 12 項目の気象要素を教師ベクトルとし、「脳梗塞データベース」の脳梗塞の搬送数を入力ベクトルとして双方向 SOM を実行した。競合層には「気象情報データベース」にある多次元の気象要素について重みベクトルが学習され、「脳梗塞データベース」の脳梗塞の搬送数も同じ競合層に学習されているため、脳梗塞の発症リスクと気象要素との関係を競合層のユニット上に視覚化することができる。それぞれのユニットには、「気象情報データベース」に学習されている気象要素と、「脳梗塞データベース」に保存されている脳梗塞の搬送数の重みベクトルが学習される。

次に、競合層に学習されている「気象情報データベース」のユニットについて、階層型クラスター分析（本研究では、ウォード法を用いた）を行った。これにより気象学的に近しいグループに分類することができ、多数の気象要素を少数の「気象パターン」に分類することができる。この気象パターンについて、気温 ( $^{\circ}\text{C}$ ) や降水量 (mm) など気象要素を比較することで、どのような「気象パターン」として説明されるかについて検討した。

さらに、一般財団法人日本気象協会が、前線の位置や気圧配置などの気象学的な基準で作成した指標で分類された天気図分類表と「気象パターン」の比較も行い、正しく気象パターンに分類されているか検証を行った。これらの検証に

より、雨天や晴天などの気象概況や、西高東低のような天気図型に分類を行った。つまり、双方向 SOM により脳梗塞の発症リスクに影響する「気象パターン」を判別した。「気象パターン」という分類指標を新たに作成できたことで、本研究で提案している双方向 SOM の有用性が確認できる。なお、自己組織化マップについての詳細な解説は Appendix に記載した。

### 3.3 決定木とサポートベクターマシン (SVM) による脳梗塞の発症予測モデル構築と予測精度の検証

ここでは、前項「3.2 双方向 SOM による気象パターンの分類」で提案している双方向 SOM により分類された「気象パターン」ごとに、発症リスクの予測モデルの構築と、このモデルの予測精度を検証する。

脳梗塞の発生頻度が多い日から「リスク・大」「リスク・中」「リスク・小」とし、対象日数を均一になるように設定した。「リスク・小」は日搬送数が3回以下とし、「リスク・大」は日搬送数が7回以上でそれ以外を「リスク・中」とした(図1)。図1では頻度分布を示しており、「リスク・小」は青色で示した箇所であり対象日数は110日間、「リスク・中」は灰色で示した箇所であり対象日数は87日間、「リスク・大」は赤色で示した箇所であり対象日数は74日間であった。このようにできるだけ均等になるように分類を行った。2002-2004年冬季の名古屋市において脳梗塞として搬送された患者数は、1435名で、1日当たり最低1名~最大13名であった。一日当たりの搬送者数の平均は5.3名であり5名が55日と最も多かった。

本研究では、脳梗塞の発症リスクを明確に比較するために「リスク・大」と「リスク・小」のみを抽出した。ここでは脳梗塞の発症リスクに与える気象要素の

特徴を確認するだけでなく、気象要素による脳梗塞の発症予測についても検討を行った。つまり、気象パターンごとに従属変数を気象要素とし、目的変数を脳梗塞の発症リスク（「リスク・大」と「リスク・小」）として、決定木とサポートベクターマシン（SVM）による分析を行った。

決定木では線形モデルの特徴から直線的な基準により脳梗塞の発症リスクを分類する（豊田，2008）。これまでも脳梗塞の発症に与え得る気象要素の影響について決定木を用いて分析した報告がある（伊藤，2006）。

SVMは非線形モデルとマージン最大化の特徴から「リスク・大」と「リスク・小」の境界を判別する（豊田，2008）。このような異なる統計手法である決定木とSVMにより、それぞれ発症リスクに与える気象要素を視覚的に表現することが可能になった。

決定木とSVMを用いて、それぞれの「気象パターン」について、発症リスクの「発症予測モデル」を構築し、それぞれ予測精度の評価を行った。さらに、双方向SOMを用いない場合の発症予測モデルを構築し、「気象パターン」に分類した場合との比較を行った。また、これらの結果から決定木学習アルゴリズムを用いた分析とSVMの予測精度について比較を行い、本研究に最適な統計手法の開発と予測精度の検討を行った。なお、決定木学習アルゴリズムを用いた分析とSVMについての詳細はAppendix（第7章）に記載した。

## 第4章 気象パターンの分類と、脳梗塞の発症リスクの結果と考察

### 4.1 はじめに

先行研究では、季節ごとに連続した長い期間における脳梗塞の発症と気象要素の与える影響について検討されている (Walach *et al.*, 2002; Inagawa *et al.*, 2002; Fisher *et al.*, 2004). また、脳梗塞の患者数と平均気温 (°C)、現地気圧 (hPa) などの特定気象要素と、脳梗塞の発症数 (発生数・患者数・搬送数) について相関分析や有意差検定、比較検定が行われてきた。このような比較的単純な統計手法による解析が多く行われてきた (Setzer *et al.*, 2007; Hong *et al.*, 2003; Woo *et al.*, 1991).

しかし、病気の発症は気温 (°C) や気圧 (hPa) など特定の気象要素が直接的に影響しているだけでない。脳梗塞の発症には、当日の気象要素だけではなく、前日の気象要素なども影響されるとの報告がある (Chen *et al.*, 1995; Connor *et al.*, 2002). このように多くの気象要素が相互に関連して複雑なメカニズムで発症に影響を与えている。先行研究では脳梗塞の発症と気象要素の関係について十分に検討されていない。

また、気象要素は前線の位置や気圧配置によって、特徴が全く異なるため、これらにも対応できる解析手法が必要となる。つまり、気象学的な特徴 (晴天・雨天・曇天・高気圧・低気圧など) によって、気象要素との関係性が全く異なっているため、これらの気象学的な特性を総合的に考慮した研究が求められる。

本研究では、双方向 SOM や階層型クラスター分析を用いて、脳梗塞に与える気象要素の影響について総合的に分析することで、脳梗塞の発症に影響を与える影響を検討した。また、決定木学習アルゴリズムを利用した方法と SVM を用



いて、この気象パターンごとに脳梗塞の発症リスクの予測モデルを構築し、予測精度の評価を行った。

## 4.2 方法

本研究では脳梗塞と、多次元で相互に関係している気象要素の関係を総合的に学習し、二次元平面上に視覚的に表現できる双方向 SOM を用いた。「脳梗塞データベース」には脳梗塞の発症数（搬送数）がデータベースとして集計されているため、入力層の入力ベクトルとした。また、解析に用いた気象情報データベースには最高気温（℃）や海面平均気圧前日差（hPa）など 12 の気象要素が多次元で存在している出力層を教師ベクトルとして、双方向 SOM により分析を行った。

これにより脳梗塞の発症数（=入力ベクトル）と、多次元に存在している気象要素（=教師ベクトル）が、双方向 SOM の競合層において 25 個のユニット（5×5 の格子状、二次元平面上）に学習され反映されている。この 25 個それぞれのユニットには教師ベクトルと入力ベクトルの重みベクトルが学習されている。つまり、双方向 SOM を用いたことで、脳梗塞の発症に影響を及ぼす気象要素の関係を二次元平面上に可視化することが可能である。

次に、競合層に学習されている 25 個のユニットについても詳細に分析を行うことで、それぞれのユニットに対応している気象要素の特徴を把握する。それぞれのユニットには解析用の気象情報データベースに含まれている全ての気象要素が重みベクターとして競合層に学習されているため、それぞれのユニットで類似した特徴を持つユニットに分類する。

本研究ではこの 25 ユニットについて「気象情報データベース」に保存されて

いる気象要素の重みベクターの階層型クラスター分析を行った。これにより類似した挙動を示すグループを分類することができる。つまり、気象パターンの分類を行った。本研究ではフリー統計ソフト「R」を用いて、Packageは“prcomp”で階層型クラスター分析（ウォード法）により分析した。

本研究では、2つの気象学的な指標に基づいて、双方向SOMと階層型クラスター分析により分類された「気象パターン」が正確にされているか検証した。

まず、気温（℃）や降水量（mm）、日照時間（h）など一般的な気象概況に直結する気象要素と比較することで、それぞれの気象パターンの特徴についての解析を行った。次に、毎日の実況天気図について気象学的な基準で分類された天気図型との比較を行うことで、双方向SOMによる気象パターンの分類結果について正確に分類されているか検討を行った。毎日の実況天気図は、一般財団法人日本気象協会東海支社が、名古屋における実況天気図について、前線の位置・気圧配置・気圧（hPa）などの気象学的な指標に基づいて作成した天気図分類フローを用いて毎日の実況天気図（名古屋を中心とした午前9時の天気図）を分類している。本研究では、北高南低（NHSL）、東高西低（EHWL）、南高東低（SHEL）、西高東低（WHEL）、高気圧部（HP）、低気圧部（LP）の6つに分類した。以下に略称を示す。

NHSL; North High and South Low pressure distribution type

EHWL; East High and West Low pressure distribution type

SHEL; South High and East Low pressure distribution type

WHEL; West High and East Low pressure distribution type

HP; High atmospheric pressure

LP; Low atmospheric pressure

このような分類を行い、日々の実況天気図を正確に表現している指標とし、

「天気図型分類表」を構築した。

このような2つの指標により、気象パターンの分類について検証した。また、競合層に学習されている「脳梗塞データベース」から、脳梗塞のリスクについても検討を行い、それぞれの気象パターンにおける発症リスクの特徴を把握した。

### 4.3 結果と考察

#### 4.3.1 気象パターンの分類

本研究では脳梗塞の発症に与える気象要素の関係を双方向 SOM と階層型クラスター分析を用いた統計解析を行い、近しい気象条件のグループに分類する手法の開発を行った。

「脳梗塞データベース」と「解析用気象情報データベース」を双方向 SOM により分析し、各ユニットに学習されている気象情報について階層型クラスター分析を行った結果、各ユニットの気象学的な特徴が明確となった3つのグループに分類できたため、これらを「気象パターン」とした(図3, 図4)。3つに分類された気象パターンについての考察を行うために、それぞれを「グループ A」、「グループ B」及び「グループ C」とした。図3, 図4から「グループ A」にはユニット番号が1, 2, 3, 4, 5, 7, 9, 「グループ B」は8, 16, 21, 22, 24, 25, 「グループ C」は6, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 23であった。グループ A は45日間, グループ B は60日間, グループ C は166日間と分類された。

それぞれの気象パターンと気象要素の検討を行なった。明確な特徴が表れた降水量(mm)と日照時間(h)について考察する。気象パターンの中で「グループ A」

にのみ 1 mm 以上の降水が観測された。また、「グループ A」は平均 10 mm の降水量があり、日照時間も平均 1.5 時間しか観測されていない。このように「グループ A」は雨天に特徴付けられるの天候である。しかし、「グループ B」と「グループ C」には明確に特徴な違いが確認できなかった（図 5）。

次に、実況天気図を気象学的な指標に基づいて分類された「天気図型分類表」と、双方向 SOM で得られた気象パターンを比較することで、双方向 SOM で気象パターンの分類が効果的にできているか検討した。表 2 のように「天気図型分類表」で分類している北高南低（NHSL）、東高西低（EHWL）、南高東低（SHEL）、西高東低（WHEL）、高気圧部（HP）、低気圧部（LP）と「グループ A, B, C」の比較を行った。

その結果、グループ A は低気圧部（LP）に 50%（23 日/45 日＝グループ A で低気圧部に分類された日数/グループ A の総数）が集中し、他のグループに比べ低気圧部にもっとも多く分類されているのはグループ A であった。これらの結果から、グループ A は“雨天”と判断した。また、グループ A に分類された日の実況天気図を参照したところ、低気圧が名古屋の上空に停滞している場合や、低気圧が名古屋に接近または離脱している途中の天気図であった。参照している実況天気図が午前 9 時の時点でのものであるため、グループ A であっても低気圧が中心にあるとは限らず、時間的に前後している場合も確認された。

グループ B も同様に天気図型分類表との比較を行った。その結果、西高東低に 81.7%（49 日/60 日＝グループ B で西高東低に分類された日数/グループ B の総数）が集中したため、“西高東低”を比較的に正しく分類できていることが確認された（表 2）。また、グループ B に分類されている日の実況天気図を参照したところ、気圧配置は“西高東低”に分類されていた。以上の結果から、グループ B は“西高東低”の気圧配置にあるケースと判断した。

グループ C も同様に天気図型分類表と比較を行った. その結果, 高気圧が存在する場合に 86.9% (53 日/61 = グループ C で高気圧が存在した日数/天気図型分類表で高気圧部の総数) が集中している (表 2).

このように双方向 SOM によって分類された気象パターンによる分類が, 天気図型分類表と一致するか検討した. また, グループ C に分類されている日の実況天気図を参照したところ, 高気圧が名古屋の上空に停滞している場合や, 高気圧が名古屋に接近または離脱している場合だけでなく, 付近に強い高気圧が複数存在している天気図が多いことが特徴が認められた. 以上の結果から, 「グループ C」は“高気圧”で特徴づけられると判断した.

これらの検証結果から, グループ A は雨天, グループ B は西高東低の気圧配置, グループ C は晴天とした. つまり, 双方向 SOM により脳梗塞の発症リスクに影響する気象要素の関係を詳細に統計解析することで, 気象パターンを正確に判別することが可能である.

双方向 SOM により脳梗塞に影響を及ぼす気象パターンという分類指標を新たに作成した. また, 本研究で提案している双方向 SOM は, 脳梗塞の発症に与える気象要素の影響を総合的に判断する際には有効な統計手法であることが確認できた.

以上の結果から日々の気象パターンを比較的, 容易に分類することができるため, 天気図型分類表などを用いずに気象概況を把握することが可能となり, 気象パターンと脳梗塞の発症リスクの関係を, 客観的な解析を行うことが可能となった. .

#### 4.3.2 脳梗塞の発症リスク

本研究では、教師ベクトルを脳梗塞データベースとし、入力ベクトルを解析用の気象情報データベースとして双方向 SOM を行っているため、競合層には脳梗塞の発症リスクと気象要素が、それぞれ重みベクトルとして双方向から学習されている。前項「4.3.1 気象パターンの分類」では、競合層の各ユニットに入力ベクトルとして学習されている解析用の気象情報データベースの気象要素について階層型クラスター分析を行い、脳梗塞の発症に影響を与える気象要素の関係を総合的に判断した結果から、気象パターンが正確に分類できていることが明らかである。

そこで、競合層に教師ベクトルとして学習されている脳梗塞データベースの発症リスクについて検討を行った。つまり、競合層の各ユニットに学習されている脳梗塞の発症リスクについて階層型クラスター分析を行い、脳梗塞の発症リスクについて分類を行った。その結果、脳梗塞の発症リスクを 2 つ（「発症リスク・大」「発症リスク・小」）に分類することに成功した。「発症リスク・大」は発症リスクが高い日とし、「発症リスク・小」は発症リスクが低い日とした（図 1）。

これまでの結果から競合層の各ユニットについて、教師ベクトルでは 3 つの気象パターンに分類ができ、入力ベクトルでは発症リスクを 2 つ（「発症リスク・大」「発症リスク・小」）に分類することができた。つまり、気象パターンごとに発症リスクを判別する手段が確立され、その結果、グループ A における「発症リスク・大」と「発症リスク・小」、グループ B における「発症リスク・大」と「発症リスク・小」、グループ C における「発症リスク・大」と「発症リスク・小」に分類することが可能となった。

このように競合層を双方向から分類した結果、それぞれの気象パターンにおける発症リスクの関係について、実況天気図と比較することが可能となった。

グループ A (雨天) において、脳梗塞の発症リスクが高い日の例は、2004 年 12 月 31 日で日搬送数は 12 件であった。この日は降水量が 11 mm であり、最低気温は 0.2°C、最高気温は 4.4°C と気温が一日中低い日であった。また、発症リスクが高い日には、寒冷前線が名古屋上空を通過しているときに発症リスクが増加している傾向が認められた。一方で、グループ A (雨天) において、脳梗塞のリスクが低い日は 2002 年 12 月 7 日であり、日搬送数は 1 件であった。この日は降水量が 4 mm、最低気温は 6.3°C、最高気温は 10.4°C であり、停滞前線が近づいていた (図 6)。このように雨天では、前線の発達している過程や名古屋からの距離、気圧配置が脳梗塞の発症に影響を及ぼしていることは明らかである。

グループ B (西高東低) において、脳梗塞の発症リスクが高い日は、天気図から前線が名古屋地域を通過した直後であったため、前日に降水の観測が多いのが特徴であった。一方で、グループ B (西高東低) において、発症リスクが低い日は、前線が通過している場合や、遠くにある場合や、高気圧が近くにあるときが多いのが特徴である (図 7)。このように西高東低では前線の移動が影響していると示唆される。西高東低の天気図は一般的な冬型の気圧配置であり、本研究対象期間の 53.5% (145/271; 研究対象期間に天気図型分類表で西高東低と分類された日数/研究対象期間) が西高東低に分類されていた。本研究から不安定な大気となっている日に脳梗塞の発症リスクが高くなる傾向が示唆される。

グループ C (晴天) において、脳梗塞の発症リスクが高い日は、高気圧が名古屋地域の近い距離に存在している日が多いのが特徴であった (図 8)。気温が高い場合や、気温日較差が 11.7°C と急激に気温が変化している日であったため、晴天では気温による影響を強いことが示唆された。この結果はこれまでも多く報

告されているが、晴天では特に気温の影響を強くうけることが明らかである。

以上のように入力ベクトルと教師ベクトルの双方向から、競合層に対して学習されている双方向 SOM の特徴により、脳梗塞の発症に影響を与える気象要素を総合的に統計処理した結果、脳梗塞の影響に関する気象パターンを正確に把握することが可能である。つまり、天気図型分類表などのフローによる天気図型の分類では把握することができない、脳梗塞の発症に特化した気象パターンの分類が可能になった。

これらの研究成果から気象パターンに分類することで、脳梗塞の発症に関連している実況天気図の特徴を正確に分類することに成功し、さらに脳梗塞の発症リスクも分類が成功している。脳梗塞の発症は天気図型によって発症に影響を与える気象要素が異なっていることが判明した。



## 第5章 発症リスクの予測精度

### 5.1 はじめに

これまでの研究（「4.3.2 脳梗塞の発症リスク」）から、双方向 SOM により同じ特徴を持つ気象パターンに分類することに成功し、また脳梗塞の発症リスクの分類にも成功した。その結果、それぞれの気象パターンにおける発症に与える気象要素やメカニズムが異なっている事実が明らかとなった。

脳梗塞の発症リスクに与える気象要素の特徴を確認するだけでなく、気象要素による脳梗塞の発症予測について検討を行った。つまり、気象パターンごとに、従属変数を 12 の気象要素とし、目的変数を脳梗塞の発症リスク（「リスク・大」と「リスク・小」とした発症予測モデルの構築を行った。ここでは、それぞれの気象パターンについて、決定木学習アルゴリズムを利用した方法とサポートベクターマシン (SVM) による分析を行い、発症予測モデルの予測精度の評価もを行い、モデルの最適化について検討をした。

### 5.2 方法

本研究では脳梗塞の発生頻度が多い日から「リスク・大」「リスク・中」「リスク・小」として対象日数を均一になるように設定する。脳梗塞の発症リスクを明確に比較するために「リスク・大」と「リスク・小」のみを抽出し解析に用いた（図 5-1）。本研究では「リスク・中」をブランクとして解析には用いない。

ここでは脳梗塞の発症リスクに与える気象要素の特徴を確認するだけでなく、気象要素による脳梗塞の発症予測についても検討を行う。つまり、「気象パターン」ごとに従属変数を気象要素とし、目的変数を脳梗塞の発症リスク（「リスク・

大」と「リスク・小」として、決定木学習アルゴリズムを利用した方法と SVM を用いて発症の予測を行った。

線形モデルである決定木学習アルゴリズムを利用した方法は直線性を基準により脳梗塞の発症リスクを分類する。このため出力結果がわかりやすく可視化できるのが特徴であり多くの研究で利用された実績がある。SVM は非線形モデルとマージン最大化の特徴から「リスク・大」と「リスク・小」の境界を判別することができ、その高い汎用性から多数の研究で利用されてきた。これらについては Appendix にて詳細な解説をしている。

発症リスクの予測精度を検証するために、決定木学習アルゴリズムを用いた方法では予測用データベースを学習用データと検証用データに分割した。学習用データは予測用データベースからランダム抽出された 70%、検証用データをそれ以外の 30%とし、本研究では第二ノードまでを有効とした。この条件下で、学習用データについて決定木決定木学習アルゴリズムを用いた方法では、学習された決定木アルゴリズムを検証用データで検証することで予測精度の検証を行った。さらに、同様の操作をランダムに抽出された学習データで決定木分析を 3 回繰り返し、それぞれ正答率を求め、この 3 回の正答率の平均値を発症の予測精度の評価に用いた。

また、SVM でも予測用データベースから発症リスクの予測精度を検証した。SVM の従属変数は、決定木決定木学習アルゴリズムを用いた方法によって第 1 ノードと第 2 ノードで分岐として使われた気象要素とした。また、決定木決定木学習アルゴリズムを用いた方法と同様に目的変数を発症リスクの指標である「リスク・大」と「リスク・小」とし、従属変数は気象要素とした。SVM は K-fold cross validation (K=3)により発症の予測精度を算出した。

本研究では決定木決定木学習アルゴリズムを用いた方法と SVM の正答率は実

測値と予測値の出力結果が、一致している場合を正解とし、全体のデータにおける実測と予測が両方とも「リスク小」か「リスク大」となっている箇所の合計の割合を正答率とした。

本研究では、脳梗塞の決定木決定木学習アルゴリズムを用いた方法と SVM により、それぞれの気象パターンにおける、発症リスクの発症予測モデルを構築し、発症予測精度の評価を行った。さらに、双方向 SOM を用いない場合（研究対象期間）の発症予測モデルを決定木学習アルゴリズムを用いた方法と SVM により構築し、「気象パターン」に分類した場合の決定木学習アルゴリズムを用いた方法と SVM との発症予測の精度を比較した。このようなアプローチで、最適な発症予測モデルの構築を行った。

## 5.3 結果

### 5.3.1 研究対象期間の発症予測精度

研究対象期間全体での発症予測精度を決定木学習アルゴリズムを用いた方法と SVM により分析した。これを、これまでに報告されている研究の一般的な事例とした。本研究では冬季（12月～翌年2月）という特定の季節に分けた季節性のみの分類を行っている。

この条件で決定木学習アルゴリズムを用いた方法と SVM によりそれぞれ発症予測モデルを構築した。また、発症予測モデルの予測精度の評価を行った。その結果、決定木学習アルゴリズムを用いた方法では 53%、SVM では 56%であった。このように、SVM の発症の予測精度は決定木学習アルゴリズムを用いた方法より良いが、どちらも正確に脳梗塞の発症リスクが分類されていない（図9）。

図9から「リスク・大」と判別されている青色のエリアと、「リスク・小」と判別されている赤色のエリアは不規則な曲線を描きながら分類はされている。しかし、「リスク・小」を示す青色のエリアには、実データの「リスク・大」と「リスク・小」が多く混在していた。一方で、「リスク・大」と判別されている赤色のエリアでは、実データでも正確に分類されているが、十分な発症予測モデルを構築できていなかったため、予測精度が低くなっている。

以上の結果から、季節性でのみの分類を行っている場合は、脳梗塞の発症リスクを予測するには十分な精度で得られないことが確認された。冬季の気象要素は複雑に相互が影響しながら変化しているため、決定木学習アルゴリズムを用いた方法やSVMでは気象要素の挙動を正確に把握することができない。

### 5.3.2 気象パターンごとの発症予測精度

これまでの研究（「5.3.1 研究対象期間の発症予測精度」）から、冬季という分類だけで決定木学習アルゴリズムを用いた方法とSVMにより構築した脳梗塞の発症予測モデルでは、複雑な気象要素に対応ができずに発症の予測が有意な精度で実施できないことが分かった。

本研究で既に有用性が明らかとなっている、気象パターンごとに発症予測モデルの構築を行った。つまり、気象パターンごとに決定木学習アルゴリズムを用いた方法とSVMを用いて、脳梗塞の発症に与える気象要素の影響を評価する発症予測モデルを構築し、その発症予測モデルの発症予測精度について検証を行った結果について以下に説明する。

## グループ A “雨天” の発症予測精度

「雨天」の発症予測精度は、決定木学習アルゴリズムを使った方法では 62.5%、SVM では 69.2%であった。SVM が決定木学習アルゴリズムを使った方法に比べて予測精度が高いことが判明した（図 10）。雨天では、日照時間前日差と最低気温を基準とした分岐となった。図 10 から SVM による判別された「リスク・大」と「リスク・小」と、実データを比較したところ正確に分類されていることは明らかである。また、学習アルゴリズムを使った方法による判別と SVM による判別を比較したところ、学習アルゴリズムを使った方法では判別が難しいエリアについても SVM では説明ができていたことが明らかとなった。

学習アルゴリズムを使った方法と SVM の結果から、日照時間前日差が前日より 6.1 時間以下のときで、最低気温が 2.8℃以上のときに「リスク小」となっている。これらの結果から前日は晴天に近い天候であった場合は、脳梗塞の発症に影響が少なかったと推察される。一方で最低気温が 2℃以下では「リスク大」となっている。

最低気温と脳卒中の関係については多くの論文で報告されている（例えば、Woo et al., 1991）。本研究からも脳梗塞の発症にも最低気温が影響していることが示唆された。このように、雨天における脳梗塞の発症リスクを視覚的に表現することが可能となった。

## グループ B “西高東低” の発症予測精度

「西高東低」では海面平均気圧と平均相対湿度前日差を基準として分岐した（図 11）。学習アルゴリズムを使った方法では予測精度は 54.5%、SVM は 62.9%

である。図 11 には、SVM は非線形モデルの特徴とマージン最大化により、脳梗塞の発症リスクを効果的に分類できていることが示されている。図 11 のように実データで「リスク・大」と「リスク・小」が接近している場合でも、境界を明確に判断している。このように西高東低では気象条件の変化が雨天や晴天よりも複雑な挙動を示しているが、SVM を用いることで高い精度で分類することができる。SVM を用いることで、決定木学習アルゴリズムを使った方法に比べて 7.4% も予測精度が向上した。

SVM の結果から海面平均気圧が下がると発症リスクが高くなることが判明した。多くの研究では脳卒中と気圧の関係についての有意義な結果が得られていないが (Jakovljević *et al.*, 1996 ; Field *et al.*, 2002), 本研究によって気圧が下がることで脳梗塞のリスクが高くなる可能性が示唆された。このように、特定の気象パターンによって発症に影響する気象要素が全く異なることが判明した。

### グループ C “晴天” の発症予測精度

「晴天」の予測精度は、決定木学習アルゴリズムを使った方法では 58.6%、SVM では 67% となった。図 11 では SVM で「リスク・小」と判別されている青色のエリアが中央付近に円形で集合し、「リスク・大」を示す赤色のエリアがそのまわりを取り囲む形成をした。このように SVM のマージン最大化の特徴を活かしたことで、決定木学習アルゴリズムを使った方法に比べて 8.4% も精度が向上した (図 12)。SVM の結果から最低気温が 2-6°C で、平均相対湿度前日差が小さい日には発症リスクが小さいことが視覚的に示される。最低気温が 2°C 以下で「リスク・大」になっている点は、「雨天」でも同様に確認されていることから、

脳梗塞の発症には最低気温が強く関係していることが示唆された。また、平均相対湿度前日差からも前日差が±10%では「リスク小」であるがそれ以上になると「リスク大」になっていることから、湿度も脳梗塞の発症に強く影響していることが示唆される。

晴天における発症予測モデルを図（図 12）によって、脳梗塞の発症リスクを視覚的に表現することが可能となった。このように、脳梗塞の発症に与える気象要素は複雑な挙動であり、また複雑に関係していることが判明した。

## 第6章 結論と課題

### 6.1 結論

脳卒中や心筋梗塞などは、気象要素や気象変化によって発症する可能性が示唆されている疾患で「気象病」と呼ばれている。また、特定の季節に多く発症したり、死亡したりする病気は「季節病」と呼ばれている。脳卒中・脳梗塞は代表的な季節病・気象病である(Kaji *et al.*, 1975)。

これまでの研究から、脳卒中・脳梗塞の発症は冬季に多く、気温が上昇するときや、気温が高いときにも発症が多いことが報告されている(Walach *et al.*, 2002; Inagawa *et al.*, 2002; Fisher *et al.*, 2004)。

しかし、これまでに気圧配置や前線の場所が脳梗塞の発症に与える影響に関して詳細な報告は行われていない。そこで、本研究では脳梗塞の発症と気象要素の関係について総合的に統計解析を行い、脳梗塞の発症に影響している気象要素や気圧配置などの気象パターンの判別や、気象パターンごとの発症リスクや、それらの実況天気図との比較による確認、脳梗塞の発症リスクを予測する発症予測モデルの構築を行った。

本研究では、脳梗塞の発症に影響する気象要素を把握するために、まず脳梗塞の搬送数を教師ベクトルとし、12の気象要素を入力ベクトルとして双方向SOMを行った。これにより教師ベクトルと入力ベクトルが競合層に反映された。この競合層のユニットに学習されている気象要素について、クラスター分析を用いて分類を行った結果、「雨天」「西高東低」「晴天」の3つの気象パターンに分類することができた。

「雨天」は降水量、日照時間との比較から判断した。また、「西高東低」「晴



天」は日本気象協会が作成している天気図型分類表との比較から、正確に分類ができていることが確認された。つまり、双方向 SOM を用いたことで、脳梗塞の発症に影響を及ぼす気象パターンの分類に成功した。

また、気象パターンだけではなく、脳梗塞の発症リスクについても分類が可能となった。そこで、各気象パターンにおいて発症リスクが高い日と低い日の特徴を把握した。

「雨天」においても発症リスクが高い日の特徴は寒冷前線が名古屋上空にある場合が多く、発症リスクが低い日は停滞前線である場合が多かった。「西高東低」では発症リスクが高い日は、前線が名古屋と通過した直後であることが多く、発症リスクが低い日は既に前線が通過して遠くにある場合や、高気圧が近くにある日が多かった。「晴天」で発症リスクが高い日は、高気圧が近くさらに気温が冬場にしては高い場合と気温日較差が 10℃以上と激しい場合という特徴が見られた。このように、双方向 SOM を用いて気象パターンに分類し、さらに発症リスクも判別ができたことで、気圧配置や前線の位置・方位についても脳梗塞の発症に影響を及ぼしていることが判明した。

次に脳梗塞の発症リスクに与える気象要素の影響を決定木学習アルゴリズムを使った方法と SVM により、それぞれ発症予測モデルを構築した。また、決定木学習アルゴリズムを使った方法と SVM のそれぞれの発症予測モデルの予測精度を比較することで有効性や判別率について評価した。

発症予測モデルは、気象パターンに分類しない場合（対象期間の全データ）と、気象パターンに分類した場合（「雨天」「西高東低」「晴天」）について比較を行った。発症リスクは、脳梗塞の搬送数が多い日を「リスク大」とし、少ない日を「リスク小」とした。まず、対象期間の全てのデータでは決定木学習アルゴリズムを使った方法と SVM のどちらも「リスク大」と「リスク小」が十分

に分類できていなかった。これは、気象条件も毎日大きく変わっていることから、決定木学習アルゴリズムを使った方法と SVM では予測ができていないと考えられる。つまり、冬季のみの季節性で発症予測モデルを構築した場合は、複雑な気象要素を十分に把握して明確な基準を徹底することが困難であることが判明した。次に、各気象パターン（「雨天」「西高東低」「晴天」）において、決定木と SVM もより脳梗塞の発症予測モデルの構築を行った。その結果、SVM を用いた場合では決定木よりも、全ての気象パターンで予測精度が良くなった。また、気象パターンに分類したことで、気象条件が近い集合となり気象要素の基準を設定しやすくなった。さらに、SVM はマージン最大化の特徴を持っており、複雑な挙動をする気象要素の特徴を捉えることができていると推察される。特に、「晴天」では、決定木などの線形回帰モデルでは発症リスクの判別は難しいが、SVM では有意義な判別が可能である。

以上の結果から、双方向 SOM による気象パターンと発症リスクの分類により、これまで考慮されてこなかった前線や気圧配置が脳梗塞の発症に与える影響についての検討を行うことができた。また、マージン最大化の特徴を活かした汎用性の高い判別ができる SVM を組み合わせることで、これまでの研究より 10% 近く発症リスクの予測精度が改善することができた。つまり、本研究で新たに開発したアプローチを採用することいで、脳梗塞の発症予測が明らかに改善される。

## 6.2 展望と課題

本研究では脳梗塞の発症に与える気象要素の影響について検討を行い、生気象学的に脳梗塞の発症を予測するモデルの開発を目的とした。そこで、脳梗塞の発生数と気象情報について総合的な統計解析を行い、脳梗塞の発症を高い精度で予測することに成功した。今後、データの信頼性や、統計学的手法、検証方法についてより厳密な評価を行っていく必要がある。

### <脳梗塞のデータベースについて>

脳梗塞の発生件数は、名古屋市消防局の救急搬送データを用いている。一般的な救急搬送データは搬送時に救命医による短時間の診察で判断しているため確定診断ではない。しかし、名古屋市消防局では病院搬送後にも病院に問い合わせをして搬送後の経過についてもデータを構成しているため、本解析で用いたデータは正しく脳梗塞の発生件数を示したデータと考えられる。

### <実況天気図との比較について>

双方向 SOM により脳梗塞の発症に影響を与える特徴的な気象パターンに分類が可能とした。この気象パターンについて検証をする際に、実況天気図は午前9時のものを使用しているため、その当日の天気図を表しているとは言えない。移動速度の速い低気圧や、前線の停滞・接近・離脱が十分に把握できていない可能性がある。しかし、本研究では一般財団法人日本気象協会が分類した天気図型分類表とも比較を行い、正確に気象パターンを分類できたことが確認され、気象パターンについては、不確かさが完全に払拭されたとはいえないが、脳梗塞の発症に影響を考察に十分だと考えられる。

#### <気象要素以外の外的環境要因による発症について>

本研究では、決定木と SVM を用いて脳梗塞の発症予測を行うために気象要素を従属変数とした。その結果、これまでよりも精度が高く脳梗塞の発症を予測することが可能となった。しかし、実際には気象要素以外の内外的な要因も、脳梗塞の発症に関係していることが考えられる。つまり、社会因子や個人の行動因子、個人の健康状態、ストレスなども発症を誘発する因子として影響していることが考えられる。今後は、個人のライフログを活用することで、予測精度を向上に貢献するかの実証は今後の課題である。本研究で用いたアプローチは、個人のライフログ等を気象データと同様に用いるものである。本研究を拡張することで、気象、社会因子や個人の行動因子、個人の健康状態、ストレスなどの指標を持ちいた脳梗塞の発症予測モデルの構築が可能である。

#### <地域性について>

本研究では名古屋市消防局が管轄している名古屋市における脳梗塞と気象要素の関係について分析を行っている。つまり、本研究では名古屋における脳梗塞の影響について研究しているため、日本全国には対応していない。この研究を別の地域で行った場合は全く異なる傾向を示す可能性が予測される。地域によって、気候や地理的環境、人口規模、緑被率などが異なっているため、発症に与える外的要因も変化している可能性がある。地域固有の予測モデルを構築していくことは、実際にこのアプローチを活用するうえで重要であると考えられる。

## 謝辞

本研究を進めるにあたり,多くの有益なご指導をいただいた指導教員の森本宏教授には深く感謝します.また,セミナーを含め日常の議論を通じて多くの知識や示唆をいただいた西田民人先生,森本研究室の二階堂暢晃氏に感謝します.また、本研究を進めるにあたって救急搬送データを提供していただいた名古屋市消防局関係者の皆様に深く感謝申し上げます.

## 引用文献

- Chen, R., Wang, C., Meng, X., Chen, H., Thach, T. Q., Wong, C. M., & Kan, H. (2013). Both low and high temperature may increase the risk of stroke mortality. *Neurology*, 81(12), 1064-1070.
- Chen, Z. Y., Chang, S. F., & Su, C. L. (1995). Weather and stroke in a subtropical area: Ilan, Taiwan. *Stroke*, 26(4), 569-572.
- Connor, M. D. (2002). Does the weather influence stroke incidence? *Stroke: a journal of cerebral circulation*, 33(7), 1751-1758.
- Cristianini, N., & Shawe-Taylor, J. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge university press.
- Feigin, V. L., Nikitin, Y. U., Bots, M. L., Vinogradova, T. E., & Grobbee, D. E. (2000). A population - based study of the associations of stroke occurrence with weather parameters in Siberia, Russia (1982-92). *European Journal of Neurology*, 7(2), 171-178.
- Field, T. S., & Hill, M. D. (2002). Weather, Chinook, and stroke occurrence. *Stroke*, 33(7), 1751-1758.

Fischer, T., Johnsen, S. O. R. P., Pedersen, L., Gaist, D., Sørensen, H. T., & Rothman, K. J. (2004). Seasonal variation in hospitalization and case fatality of subarachnoid hemorrhage—a nationwide Danish study on 9,367 patients. *Neuroepidemiology*, 24(1-2), 32-37.

Goggins, W. B., Woo, J., Ho, S., Chan, E. Y., & Chau, P. H. (2012). Weather, season, and daily stroke admissions in Hong Kong. *International journal of biometeorology*, 56(5), 865-872.

Hong, Y. C., Rha, J. H., Lee, J. T., Ha, E. H., Kwon, H. J., & Kim, H. O. (2003). Ischemic stroke associated with decrease in temperature. *Epidemiology*, 14(4), 473-478.

Inagawa, T. (2002). Seasonal variation in the incidence of aneurysmal subarachnoid hemorrhage in hospital-and community-based studies. *Journal of neurosurgery*, 96(3), 497-509.

Jakovljević, D., Salomaa, V., Sivenius, J., Tamminen, M., Sarti, C., Salmi, K., ... & Tuomilehto, J. (1996). Seasonal Variation in the Occurrence of Stroke in a Finnish Adult Population The FINMONICA Stroke Register. *Stroke*, 27(10), 1774-1779.

Sato, J., Morimae, H., Seino, Y., Kobayashi, T., Suzuki, N., & Mizumura, K. (1999). Lowering barometric pressure aggravates mechanical

- allodynia and hyperalgesia in a rat model of neuropathic pain. *Neuroscience letters*, 266(1), 21-24.
- Kawai, K., Nonaka, K., Suzuki, H., Kirino, T., & Tamura, A. (2001). Differential effects of activity and climate on onset of subarachnoid hemorrhage. *Neurologia medico-chirurgica*, 41(5), 229-36.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological cybernetics*, 43(1), 59-69.
- Kovats, R. S., & Kristie, L. E. (2006). Heatwaves and public health in Europe. *The European Journal of Public Health*, 16(6), 592-599.
- Lanska, D. J., & Hoffmann, R. G. (1999). Seasonal variation in stroke mortality rates. *Neurology*, 52(5), 984-984.
- Lejeune, J. P., Vinchon, M., Amouyel, P., Escartin, T., Escartin, D., & Christiaens, J. L. (1994). Association of occurrence of aneurysmal bleeding with meteorologic variations in the north of France. *Stroke; a journal of cerebral circulation*, 25(2), 338-341.
- Liang, W., Hu, Y., Kasabov, N., & Feigin, V. (2011, January). Exploring associations between changes in ambient temperature and stroke occurrence: comparative analysis using global and personalised



modelling approaches. In *Neural Information Processing* (pp. 129-137). Springer Berlin Heidelberg.

Lim, Y. H., Kim, H., & Hong, Y. C. (2013). Variation in mortality of ischemic and hemorrhagic strokes in relation to high temperature. *International journal of biometeorology*, 57(1), 145-153.

Martens, P., Kovats, R. S., Nijhof, S., De Vries, P., Livermore, M. T. J., Bradley, D. J., ... & McMichael, A. J. (1999). Climate change and future populations at risk of malaria. *Global Environmental Change*, 9, S89-S107.

Oberg, A. L., Ferguson, J. A., McIntyre, L. M., & Horner, R. D. (2000). Incidence of stroke and season of the year: evidence of an association. *American journal of epidemiology*, 152(6), 558-564.

Ono, Y., N. Aoki, H. Horibe, N. Hayakawa and. H. Okada(1974)Biometeorologic studies on cerebrovascular diseases. V. A multivariate analysis of meteorologic effects on cerebrovascular accident. *Jpn Circ J*, 38: 195-208.

Pascal, M., Laaidi, K., Ledrans, M., Baffert, E., Caserio-Schönemann, C., Le Tertre, A., ... & Empereur-Bissonnet, P. (2006). France's heat health watch warning system. *International Journal of Biometeorology*, 50(3),

144-153.

Passero, S., Reale, F., Ciacci, G., & Zei, E. (2000). Differing temporal patterns of onset in subgroups of patients with intracerebral hemorrhage. *Stroke*, 31(7), 1538-1544.

Setzer, M., Beck, J., Hermann, E., Raabe, A., Seifert, V., Vatter, H., & Marquardt, G. (2007). The influence of barometric pressure changes and standard meteorological variables on the occurrence and clinical features of subarachnoid hemorrhage. *Surgical neurology*, 67(3), 264-272.

Shinkawa, A., Ueda, K., Hasuo, Y., Kiyohara, Y., & Fujishima, M. (1990). Seasonal variation in stroke incidence in Hisayama, Japan. *Stroke*, 21(9), 1262-1267.

Tripathi, S., Srinivas, V. V., & Nanjundiah, R. S. (2006). Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of Hydrology*, 330(3), 621-640.

Walach, H., Schweickhardt, A., & Bucher, K. (2002). [Does weather modify headaches? An empirical evaluation of bio-weather categorization]. *Schmerz (Berlin, Germany)*, 16(1), 1-8.

Wang, H., Matsumura, M., Kakehashi, M., & Eboshida, A (2006). Effects of

atmospheric temperature and pressure on the occurrence of acute myocardial infarction in Hiroshima City, Japan. *Hiroshima journal of medical sciences*, 55(2), 45-51.

Wang, Y., Levi, C. R., Attia, J. R., D'Este, C. A., Spratt, N., & Fisher, J. (2003). Seasonal Variation in Stroke in the Hunter Region, Australia A 5-Year Hospital-Based Study, 1995-2000. *Stroke*, 34(5), 1144-1150.

Woo, J., Kay, R., & Nicholls, M. G. (1991). Environmental temperature and stroke in a subtropical climate. *Neuroepidemiology*, 10(5-6), 260-265.

伊藤庸一郎, 佐藤省三, 須藤千春, & 堀越哲美. (2006). データマイニング手法を用いた気象と発症の関係の定量化法の検討. *人間と生活環境*, 13(1), 19-27.

加地正郎. (1975). 気象と病気. 加地正郎 (編著), 人間・気象・病気, NHK 出版協会, 東京, 159-235.

花城和彦, 玉城昇, 小杉忠誠, 嘉数朝一, 兼島洋, & 斎藤厚. (1998). 沖縄県地方の気象因子と喘息発作誘発との関連. *アレルギー*, 47(4), 434-448.

小野雅司. (2009). 地球温暖化と熱中症. *地球環境*, 14(2), 263-270.

上坂一譲, & 水谷章夫. (2003). 41213 気象および環境要因が家庭内の転落・転

倒および溺水事故に及ぼす影響（健康害，環境工学 II）. 学術講演梗概集.  
D-2, 環境工学 II, 2003, 447-448.

長谷川房雄. (1985). 脳卒中の発症と住環境との関係についての山形県郡部を対象とした調査研究. 日本公衆衛生雑誌, 32(4), p181-193.

福岡義隆. (2003). 気象・季節の感情障害への影響（特集：気象・季節と疾病）. 地球環境, 8(2), 221-228.

豊田秀樹. (2008). データマイニング入門: R で学ぶ最新データ解析. 東京図書.

豊田章宏. (2011). 全国労災病院 46,000 例からみた脳卒中発症の季節性 (2002 - 2008 年). 脳卒中, 33(2), 226-235.

汪宏莉, 梯正之, 松村誠, & 烏帽子田彰. (2007). 急性心筋梗塞の発症と気象条件の関連性について. J Cardiol, 49(1), 31-40.

総務省消防庁ホームページ:平成 24 年版消防白書 第 2 章消防防災の組織と活動. 第 5 節 救急体制. 5. 救急業務を取り巻く課題 (7) 救急需要増への対応.

## Appendix

### 第7章 解析手法

#### 7.1 SOM (自己組織化マップ: Self-Organizing Map)

##### 7.1.1 SOM

SOM (Self-Organizing Map) はニューラルネットの一種であり, 多次元データを二次元格子状に投射し, 可視化することによって傾向分析や予測を行う手法である (Kohonen *et al.*, 1982; Kohonen *et al.*, 1990; Wehrens *et al.*, 2007). SOM は工学・医学・気象学など, 様々な分野で用いられている (Liu and Weisberg *et al.* 2011; Leloup *et al.* 2007; Tozuka *et al.* 2008; Sugimoto *et al.* 2008; Sakai *et al.* 2010).

SOM は入力層と競合層の二層により構成されている. 入力層には多数の入力ベクトルの集合が存在し, 個々の入力ベクトル  $x_k$  は多次元のデータをもっている ( $x_k \in \mathfrak{R}^N$ ). 入力ベクトルの数を  $K$  とすると入力ベクトルの集合 ( $X$ ) は次式で表すことができる.

$$X = \{x_1, x_2, \dots, x_k, \dots, x_K\} \quad (1)$$

競合層には, 二次元格子状に複数のユニットが配置されており, 個々のユニットには入力ベクトルと同次元の重みベクトル ( $w_i \in \mathfrak{R}^N$ ) が存在している. 競合層のユニット数を  $I$  とすると, ユニットの集合 ( $U$ ) および重みベクトルの集合 ( $W$ ) は以下のように表すことができる.

$$U = \{u_1, u_2, \dots, u_i, \dots, u_I\} \quad (2)$$

$$W = \{w_1, w_2, \dots, w_i, \dots, w_I\} \quad (3)$$

SOM の学習アルゴリズムでは、まず、 $t$  回目の学習用の入力ベクトルとして、入力ベクトルの集合から任意の入力ベクトルを選択する。 $t$  回目の学習用の入力ベクトルは次式で表すことができる。

$$x(t) = x_k \quad (4)$$

ここで、学習を終了する繰り返し数を  $T$  とすると、

$$t = 0, 1, 2, \dots, T$$

となる。

次に、ユニット間の競争により、勝者ユニットを決定する。勝者ユニットは入力ベクトルと最も類似度の高い重みベクトルを持つユニットであり、類似度は入力ベクトルと各ユニットの重みベクトルのユークリッド距離によって決定される。勝者ユニット ( $u_c$ ) の重みベクトル ( $w_c$ ) は次式を満たす。

$$\|x(t) - w_c(t)\| = \min_i \{\|x(t) - w_i(t)\|\} \quad (5)$$

次に、勝者ユニットは自らの重みベクトルを、入力ベクトルとの類似度を高めるように更新することで、入力ベクトルの特徴を学習する。また、勝者ユニットの近傍領域内のユニットも、勝者ユニットからの距離に応じて学習を行う。これらの学習は次式によって表される。

$$w_i(t+1) = w_i(t) + h_{ci}(t) \{x(t) - w_i(t)\} \quad u_c \in N_c(t) \quad (6)$$

ここで、 $N_c(t)$  は勝者ユニットの近傍領域を表す。 $h_{ci}(t)$  は近傍関数であり、次式で

表される.

$$h_{ci}(t) = \alpha(t) \exp\left(-\frac{\|r_c(t) - r_i\|^2}{2\sigma^2(t)}\right) \quad (7)$$

ここで,  $\alpha(t)$  は学習の程度を決定する学習率係数,  $\sigma(t)$  は近傍領域の大きさを決定する. また,  $r_c(t)$  および  $r_i$  はそれぞれ勝者ユニットと各ユニットの二次元平面上の位置ベクトルを表す.

この学習アルゴリズムを  $t$  が, 0 から  $T$  に達するまで繰り返す.

### 7.1.2 双方向 SOM

双方向 SOM は, 入力層と競合層に, 出力層を加えた三層からなる (図 13). 入力層には SOM と同様に多数の入力ベクトルの集合が存在する. 出力層には, 多数の教師ベクトルの集合が存在し, 個々の教師ベクトル  $y_k$  は多次元のデータを持つ ( $y_k \in \mathfrak{R}^M$ ). 教師ベクトルと入力ベクトルは同数存在し, 一対一に対応している. 教師ベクトルの集合 ( $Y$ ) は次式で表される.

$$Y = \{y_1, y_2, \dots, y_k, \dots, y_K\} \quad (8)$$

競合層には, SOM と同様に二次元格子状に複数のユニットが配置されており, 個々のユニットには入力ベクトルと同次元の重みベクトルに加えて, 教師ベクトルと同次元の重みベクトル ( $v_i \in \mathfrak{R}^M$ ) が存在する. 教師ベクトルと同次元の重みベクトルの集合 ( $V$ ) は以下のように表される.

$$V = \{v_1, v_2, \dots, v_i, \dots, v_I\} \quad (9)$$

双方向 SOM の学習アルゴリズムでは、まず、 $t$  回目の学習用の入力・教師ベクトルとして、入力・教師ベクトルの集合から任意の入力・教師ベクトルの組を選択する。 $t$  回目の学習用の入力・教師ベクトルの組は次式で表される。

$$(x(t), y(t)) = (x_k, y_k) \quad t=0,1,2,\dots,T \quad (10)$$

次に、勝者ユニットの決定を行う。双方向 SOM では類似度に、入力ベクトルと教師ベクトルのそれぞれに対する重みベクトルのユークリッド距離を考慮する。したがって、勝者ユニット  $u_c$  の重みベクトルの組  $(w_c(t), v_c(t))$  は、次式を満たす。

$$\|x(t) - w_c(t)\| + \|y(t) - v_c(t)\| = \min_i \{ \|x(t) - w_i(t)\| + \|y(t) - v_i(t)\| \} \quad (11)$$

続いて、重みベクトルの更新を行う。双方向 SOM では、二種類の重みベクトルそれぞれについて更新を行う。ユニット  $u_{ij}$  の時刻  $t+1$  における重みベクトル  $(w_{ij}(t+1), v_{ij}(t+1))$  は、時刻  $t$  における重みベクトル  $(w_{ij}(t), v_{ij}(t))$  を用いてそれぞれ次のように表される。

$$w_i(t+1) = w_i(t) + h_{ci}(t) \{ x(t) - w_i(t) \} \quad u_i \in N_c(t) \quad (12)$$

$$v_i(t+1) = v_i(t) + h_{ci}(t) \{ y(t) - v_i(t) \} \quad u_i \in N_c(t) \quad (13)$$

以上の学習アルゴリズムを  $t$  が、0 から  $T$  になるまで繰り返す。これにより、入力ベクトルと教師ベクトルの特徴が重みベクトルに反映され、二次元平面内のユニットに配置される。各ユニットに配置された二種類の重みベクトルが、双方向 SOM の最終的な出力となる。



## 7.2 決定木学習アルゴリズムを使った方法

決定木学習アルゴリズムを使った方法は、機械学習において一般的な統計手法の一種であり、予測や分類を行う際に多く用いられている。決定木学習アルゴリズムを使った方法はデータの属性を、木構造を用いて分類する手法である。決定木学習アルゴリズムを使った方法では、機械学習のアルゴリズムによって予測変数の中から重要な変数を判別して抽出し、基準変数を予測するためのルールを作成し、決定木モデルとして表される。全体集合に対して予測変数を用いて段階的にグループ分けすることで要素を分類していくため、分類の手順を理解しやすいというメリットがある（図 14）。この結果、高次元の判別を繰り返すことで、全体集合に対しての判別結果を視覚的に確認できる。

本研究では、フリー統計ソフト「R」のパッケージ”mvpart”を用いて、関数mvpartにより決定木学習アルゴリズムを使った方法による分析を行った。本研究では決定木学習アルゴリズムを使った方法の分岐基準はCARTを用いた。CARTは、分岐基準となる予測変数を選択するための評価指標としてジニ係数を用いている。つまり、最適な分岐基準をジニ係数により評価をすることで、全体集合の判別をより良い結果に導いている。

以下で、分岐の評価基準について説明する。

決定木学習アルゴリズムを使った方法は木構造で構成されており、全体集合を2つに分岐を繰り返し替えして分類する。この分岐する変数はノード(節)と呼ばれ、そのノードにより2分岐された左右の集合は枝と呼ばれている。

この分岐基準は、決定木学習アルゴリズムを使った方法において分岐される直前に集合である親ノードと子ノードの間で不純度 (impurity) という指標をもとに分岐する変数が選択される。この不純度は分岐する前後について誤差が

改善される際の指標として次の式で定義されている。

$$\delta GI(t) = P_t GI(t) - P_L GI(t_L) - P_R GI(t_R) \quad (14)$$

式(14)の  $GI(t)$  におけるノード  $t$  における Gini 係数である。

$$GI(t) = 1 - \sum_k P(k|t)^2 \quad (15)$$

式(14)のにある  $P(k|t)$  は、ノード  $t$  内のカテゴリ  $k$  が正しく判別されているかの指標となる比率を示す。また、式(14)及び(15)の各項は以下の通りである。

$GI(t_L)$  はノード  $t$  により分岐する際に左に分類された枝の Gini 係数

$GI(t_R)$  はノード  $t$  により分岐する際に右に分類された枝の Gini 係数

$P_t$  は分割する前における集合の比率

$P_L$  は分割した後の”右”における集合の比率

$P_R$  は分割した後の”左”における集合の比率

このように CART では決定木学習アルゴリズムを使った方法を利用する際に選択している従属変数について、このような不純度が最も高くなる変数を選択することで、最適な分岐を繰り返し行うことで木と枝に分岐して木構造に構築することいができる。以上のように、決定木学習アルゴリズムを使った方法は全体集合を繰り返すことで分類能が高くなる場合や、変数を増やすことで分岐する変数の選択枝を増やすことになり、分類の精度はどんどんと高くすることができる。しかし、このような決定木学習アルゴリズムを使った方法でつくられた予測モデルや分類モデルでは、分岐基準には差がない場合でも分岐を繰り返しているため分析に用いていないデータを当てはめた場合に正確な分類が行われない

可能性が高くなる.このような現象を過学習といわれ,汎用性が低くなり正しい分析結果と言えない.そこで,一般的に決定木ではこのような過学習を起こさないために,一定の基準で過学習しているノードを剪定する.作業を行うことで,決定木モデルが学習の際に用いたデータ(学習データ)に依存しすぎないようにすることができる.

本研究では,フリー統計ソフト「R」のパッケージ“mvpart”に含まれる関数 `prune` により剪定基準を設定した.これは,分岐の数と決定木の複雑さを比較した `cp`(complexity parameter)値を設定する.このCPは複雑さが収束し始めているところを剪定の基準にしている場合が多い.このように剪定基準により,剪定を行うことで過学習を起こさない適切な分類が可能となる.

また,決定木学習アルゴリズムを使った方法では精度を評価や,信ぴょう性を確認するためにデータ全体を「学習データ」「テストデータ」に2分割する.学習データのみを用いて決定木学習アルゴリズムを使った方法による分析を行い,構築された決定木モデルについて学習に全く用いられていないテストデータに対して分析を行う.これにより,正しく判別しているか確認を行うことができる.

### 7.3 SVM (サポートベクターマシン: Support Vector Machine)

SVM (Support Vector Machine) は要素集合の2つのクラスへの分類に用いられる有効な識別器である。SVM は学習データの中で識別境界の両側の線状の点がサポートベクトル (Support vector) を基準として、そのユークリッド距離が最も大きくなるような位置に識別境界を設定する (図 15)。つまり、サポートベクトルと識別境界との距離であるマージン (Margin) を最大化するように超空間平面 (Hyper-plane) を構築しクラス分類を行う。SVM の特徴はマージン最大化による汎化能力の向上と、カーネルトリック関数による非線形判別であり、分類結果を視覚的に容易に捉えることができる。本研究では代表的な RBF カーネルトリックのカーネル関数を用いた。

データ集合  $x_i = (x_{i1}, x_{i2}, \dots, x_{iN})^T \in \mathcal{R}^N, (i=1, \dots, n)$  がクラス  $t_i \in \{+1, -1\}$  に属している場合、その判別関数を  $w, h$  を用いて以下のように表す。

$$f(x_i) = w^T x_i - h \quad (16)$$

$f(x_i) = 0$  を満たすポイントのは  $N-1$  の多次元な超平面となる。この超平面 (Hyper-plane) に最も接近するデータをサポートベクトル (Support Vector) と呼ぶ。このサポートベクトルと超平面 (Hyper-plane) までの距離をマージン (Margin) という。このマージンが  $1/\|w\|$  となる。このマージンを最大化できるような  $w$  が選択される。

SVM は以下の式により非線形分離も可能となる。

$$L_D(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j t_i t_j \phi^T(x_i) \phi(x_j) \quad (17)$$

非線形型変換 $\phi$ を用いて高次元空間で線形識別を行う。実質的な非線形識別が元のデータ $x_i$ に対し、非線形変換を行った高次元特徴空間でのデータ $\phi(x_i)$ を識別する。今回は最適な識別関数を構成するためにカーネルトリックという手法を用いる。カーネルトリックとは変換前のデータを用いた関数 $K(x_i, x_j)$ により計算した。

$$\phi^T(x_i)\phi(x_j) = K(x_i, x_j) \quad (18)$$

このような関数 $K$ をカーネルという。また、本研究では最も一般的な RBF カーネルを用いた。以下に式を示す。

$$K(x_i, x_j) = \exp\left(-\gamma\|x_i - x_j\|^2\right) \quad (19)$$

このようにサポートベクターマシンはマージン最大化という特徴を用いることで、学習データに依存しすぎず未知のデータに対する汎化能力を高めることができる。

次に重交差妥当性は、グリッドサーチによって求めた最適なチューニング母数を、未知のデータに対しても妥当なものにするために行う。この方法では学習データを $k$ 群に分割し、そのうちの1群を検証データ、残りの $k-1$ 群を学習データとして全 $k$ 回交差妥当化検証を行う。結果として平均的に高い判別精度が得られたならば、そのモデルは未知のデータに対しても有意な意味を持つといえる。

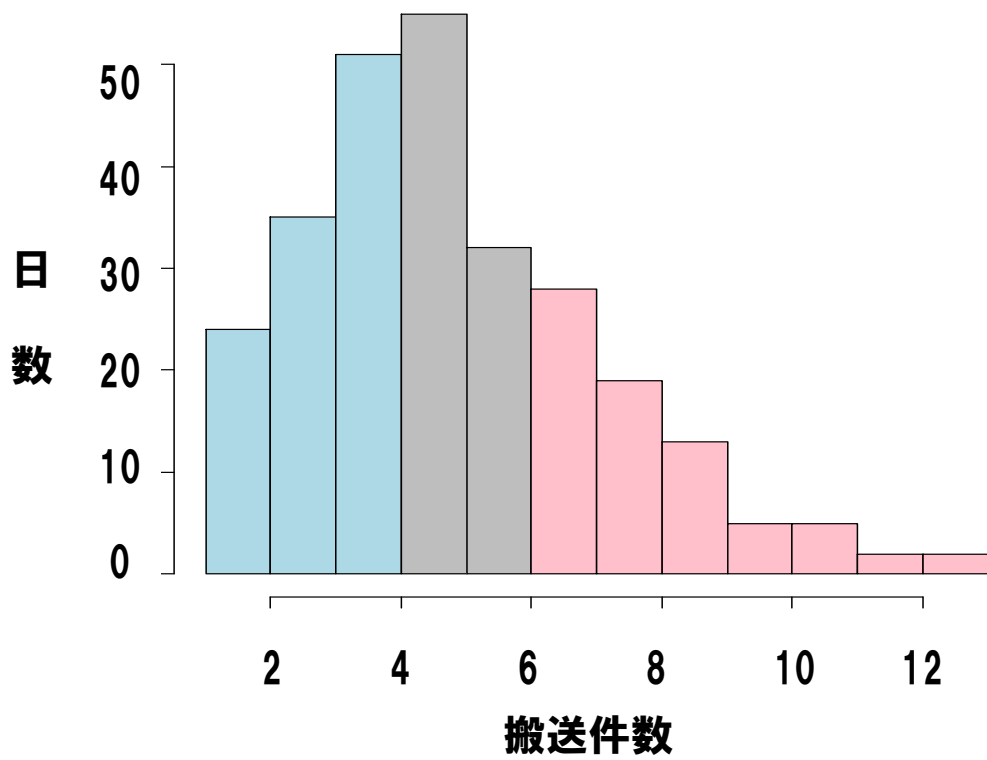


図1 2002-2004年の名古屋市における一日あたりの脳梗塞患者の救急搬送数のヒストグラム

青色：搬送数が少ない（3件以下）「リスク小」

赤色：搬送数が多い（6件以上）「リスク大」

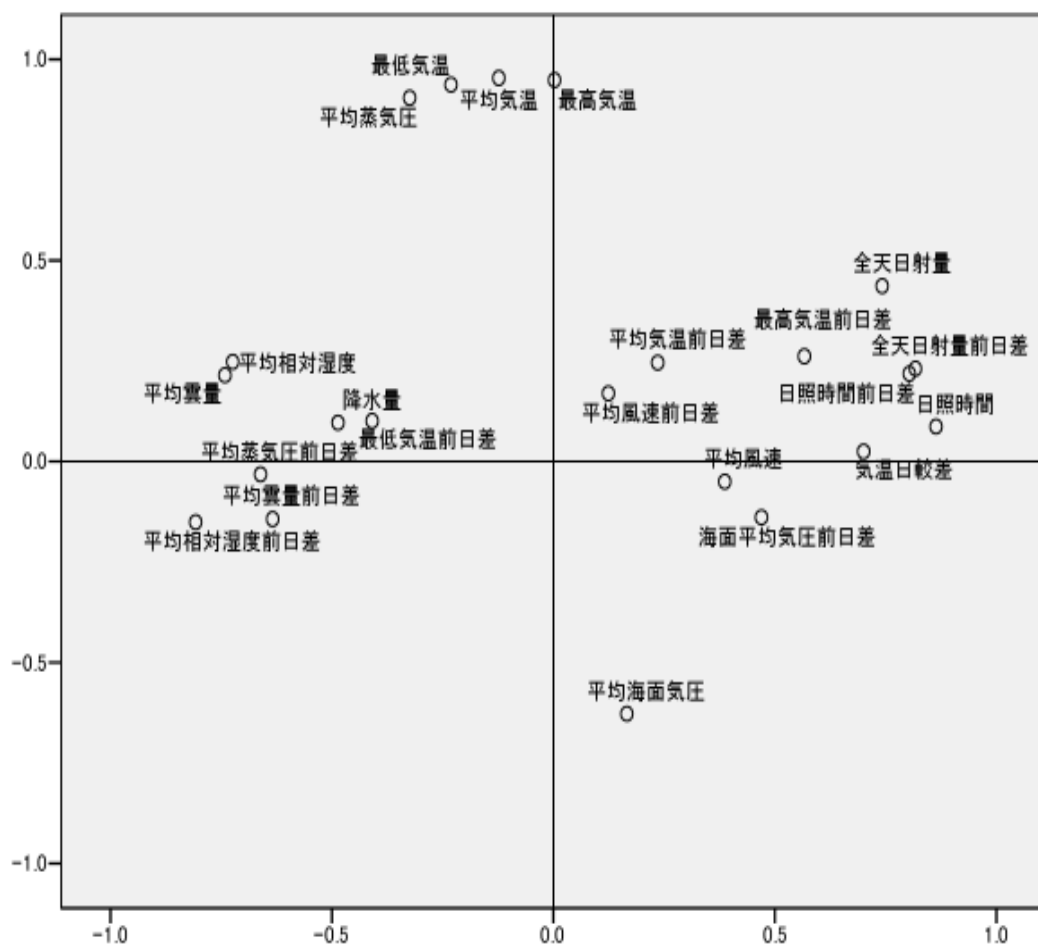


図2 主成分分析による気象要素の多重共線性回避の結果

解析用気象情報データベースとして、以下の12要素を用いる。平均海面気圧、平均気温、最高気温、気温日較差、平均相対湿度、日照時間、全天日射量、降水量、海面平均気温前日差、最高気温前日差、平均相対湿度前日差、日照時間前日差。

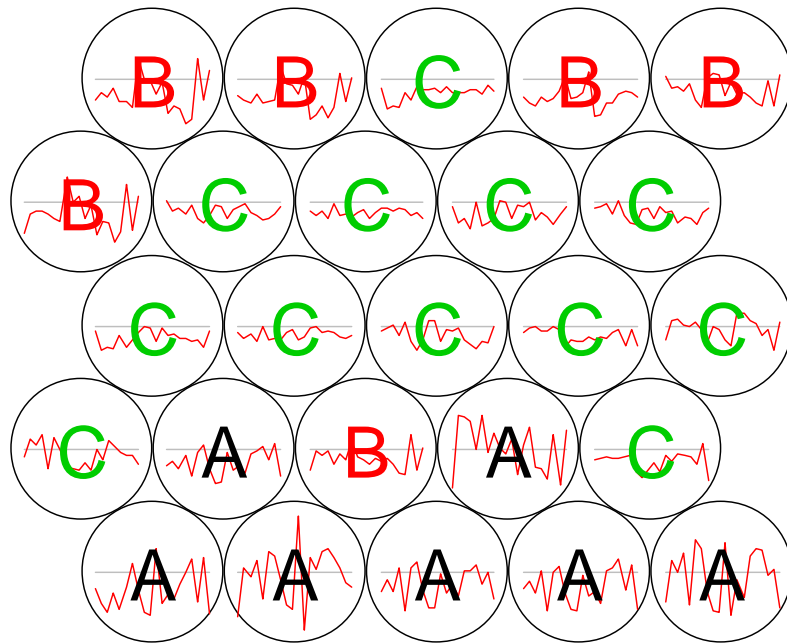


図3 双方向 SOM で学習された競合層の 25 ユニット (5×5) の 教師ベクトルである気象要素について、階層型クラスター分析を行った。

最下段の左端がユニット番号 1 となり、右端がユニット番号 5 である。2 段目は左端からユニット番号 6 で、2 段目の右端はユニット番号 10 となっている。3 段目は左端からユニット番号 11 で、4 段目の右端はユニット番号 16 となっている。5 段目は左端からユニット番号 21 で、2 段目の右端はユニット番号 25 となっている。



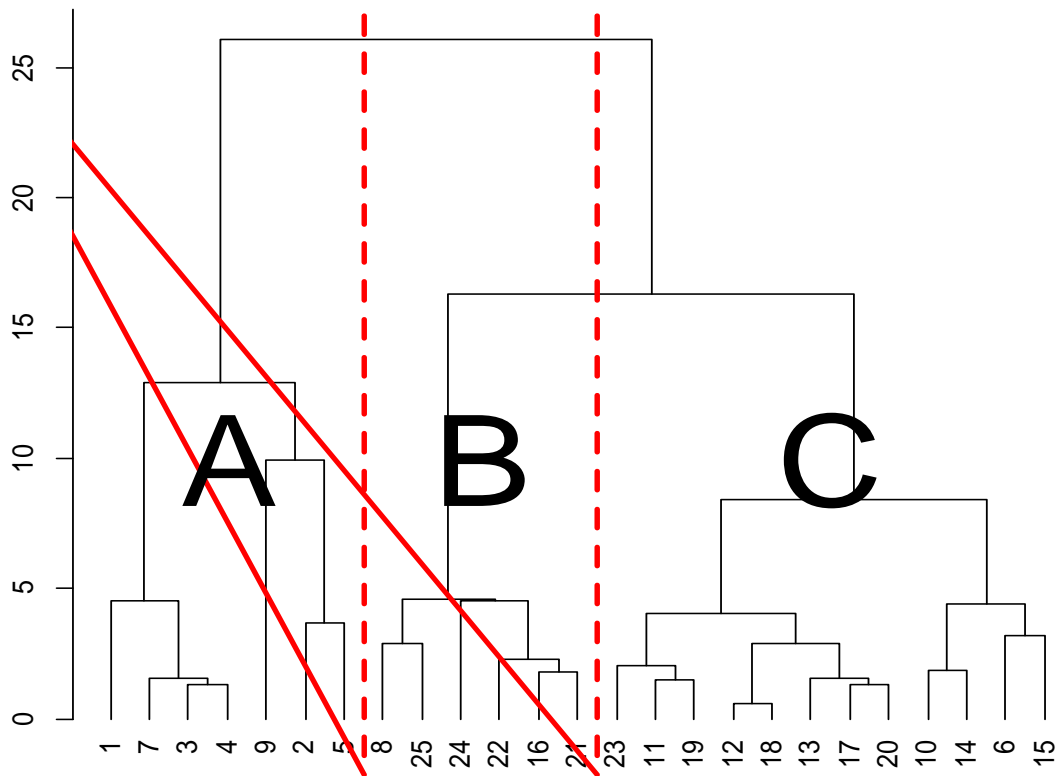
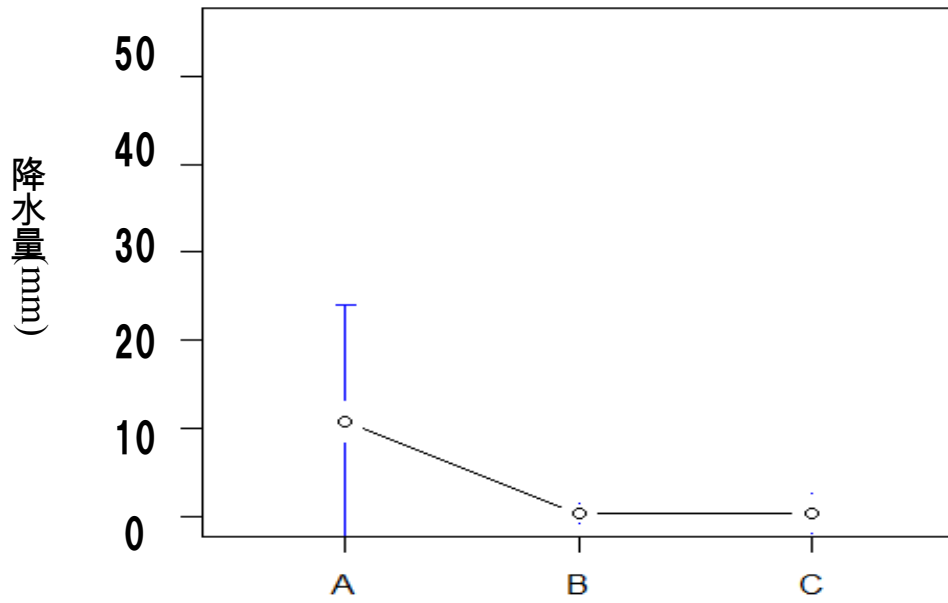
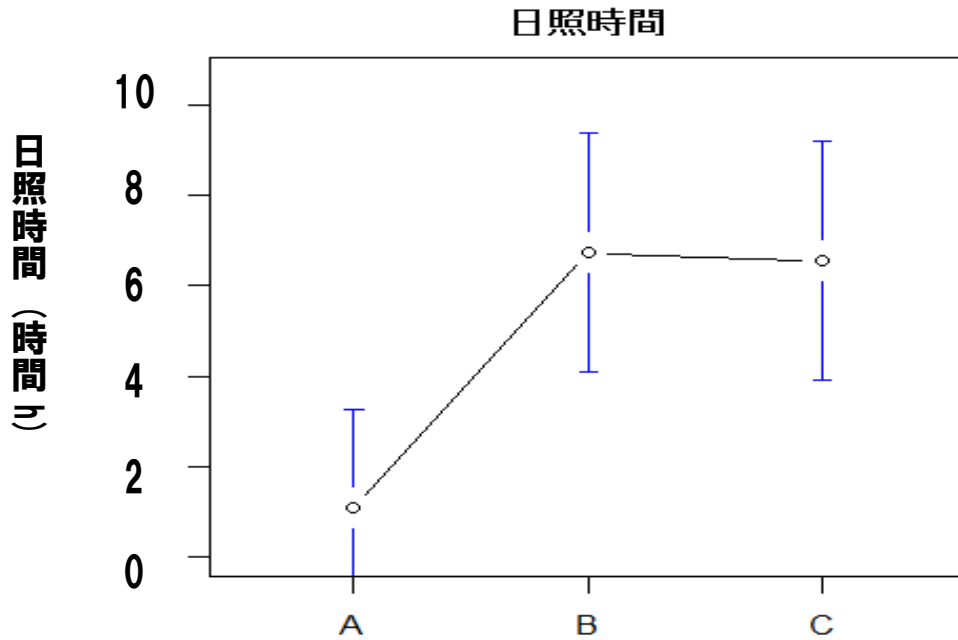


図4 双方向 SOM で学習された競合層の 25 ユニット (5×5) の教師ベクトルである気象要素について、階層型クラスター分析を行った



気候パターン



気候パターン

図 5 気候パターン (グループ A, B, C) と降水量・日照時間の比較

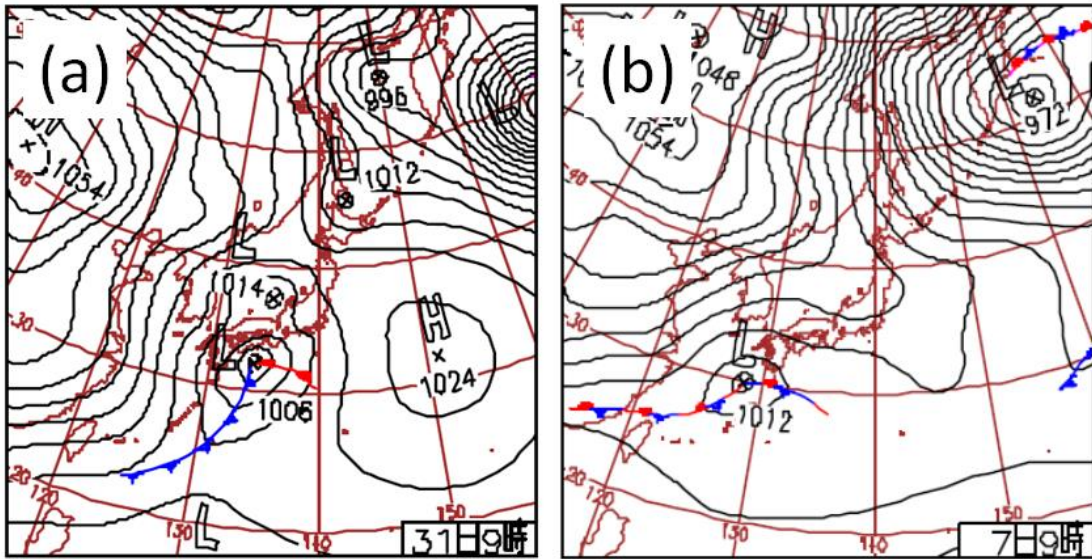


図6 グループ A「雨天」と分類された天気図

- (a) 脳梗塞の発症リスクが高い天気図 (2004年12月31日)
- (b) 脳梗塞の発症リスクが低い天気図 (2002年12月7日)

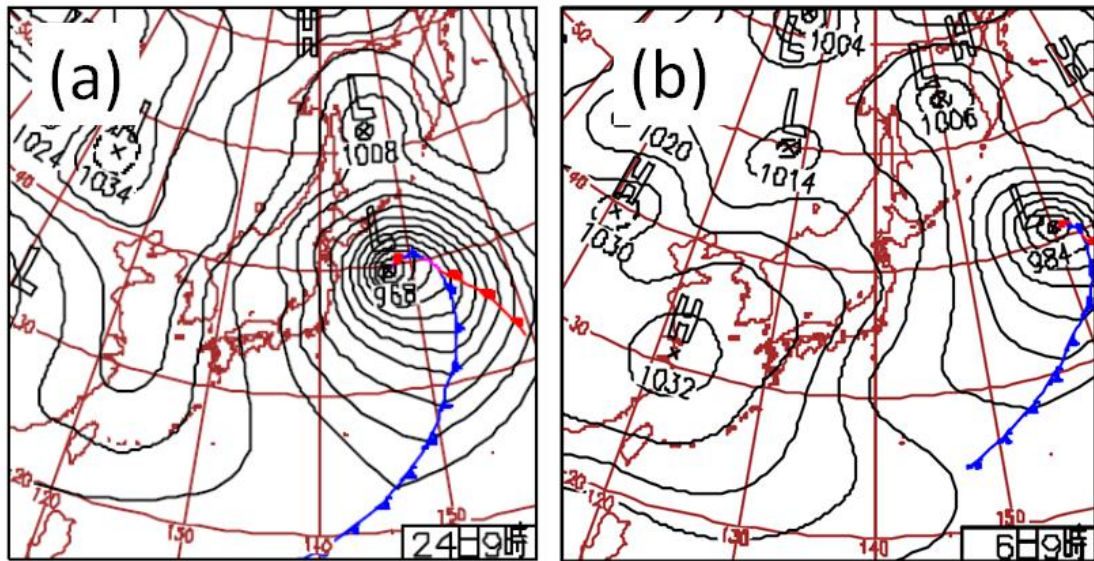


図7 グループB「西高東低」に分類された天気図  
 (a) 脳梗塞の発症リスクが高い天気図 (2003年12月24日)  
 (b) 脳梗塞の発症リスクが低い天気図 (2003年2月6日)

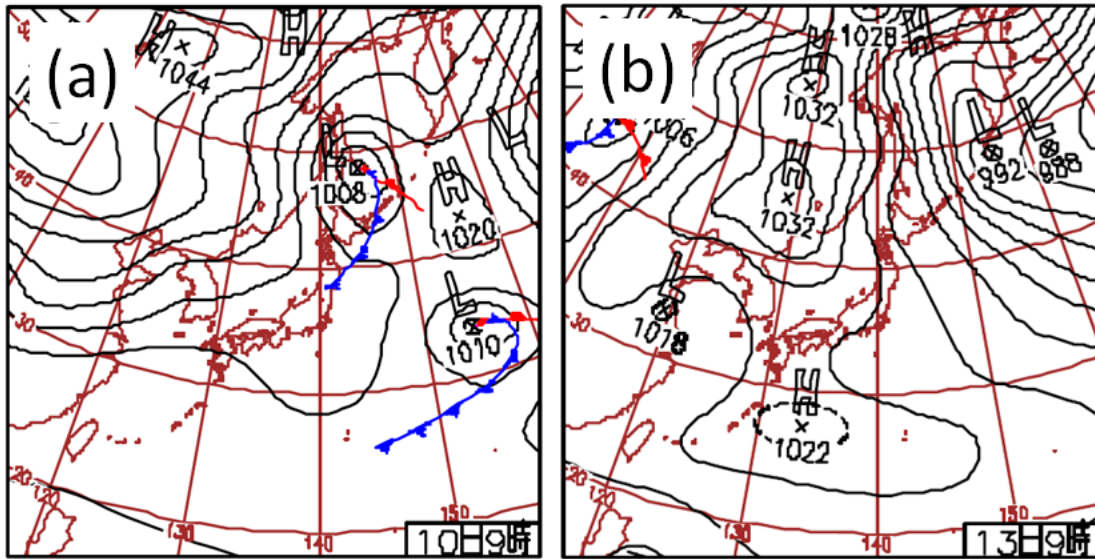


図8 グループC「晴天」に分類された天気図

- (a) 脳梗塞の発症リスクが高い天気図 (2003年2月10日)
- (b) 脳梗塞の発症リスクが低い天気図 (2003年1月13日)

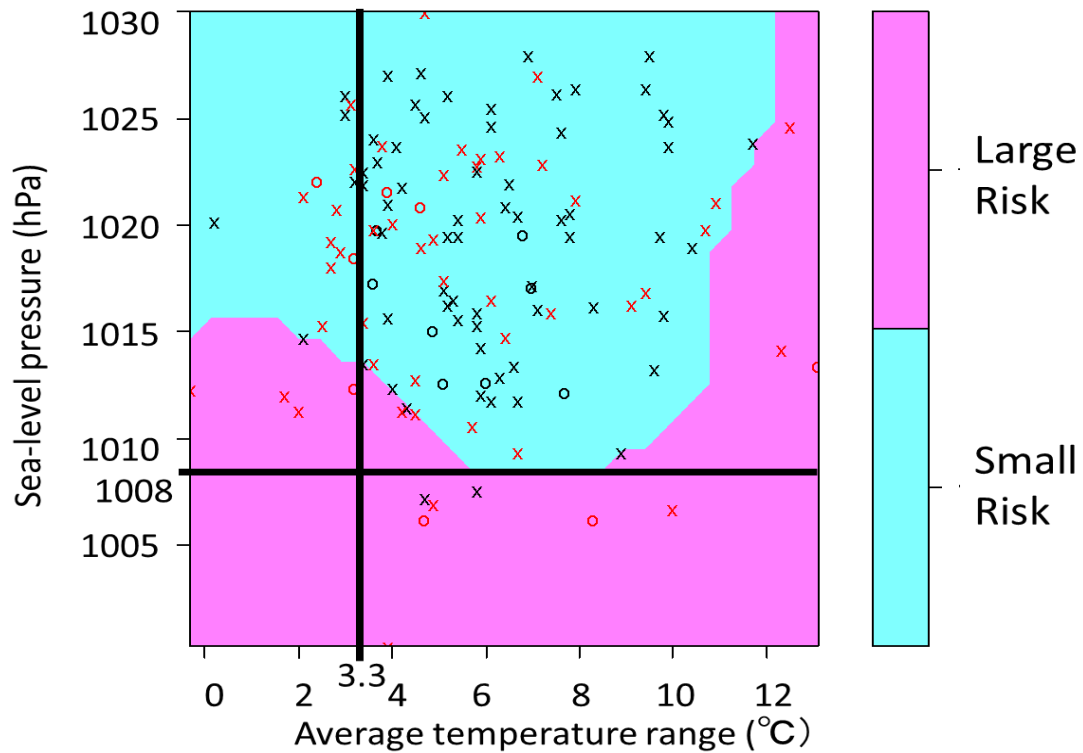


図9 対象期間における SVM と決定木の比較

○：サポートベクター

×：実データ

(赤：「リスク・大 (Large Risk)」, 黒；「リスク・小 (Large Small)」)

赤色のエリア：SVMにより「リスク・大」と判別されたエリア

青色のエリア：SVMにより「リスク・小」と判別されたエリア

実線：決定木の分岐基準

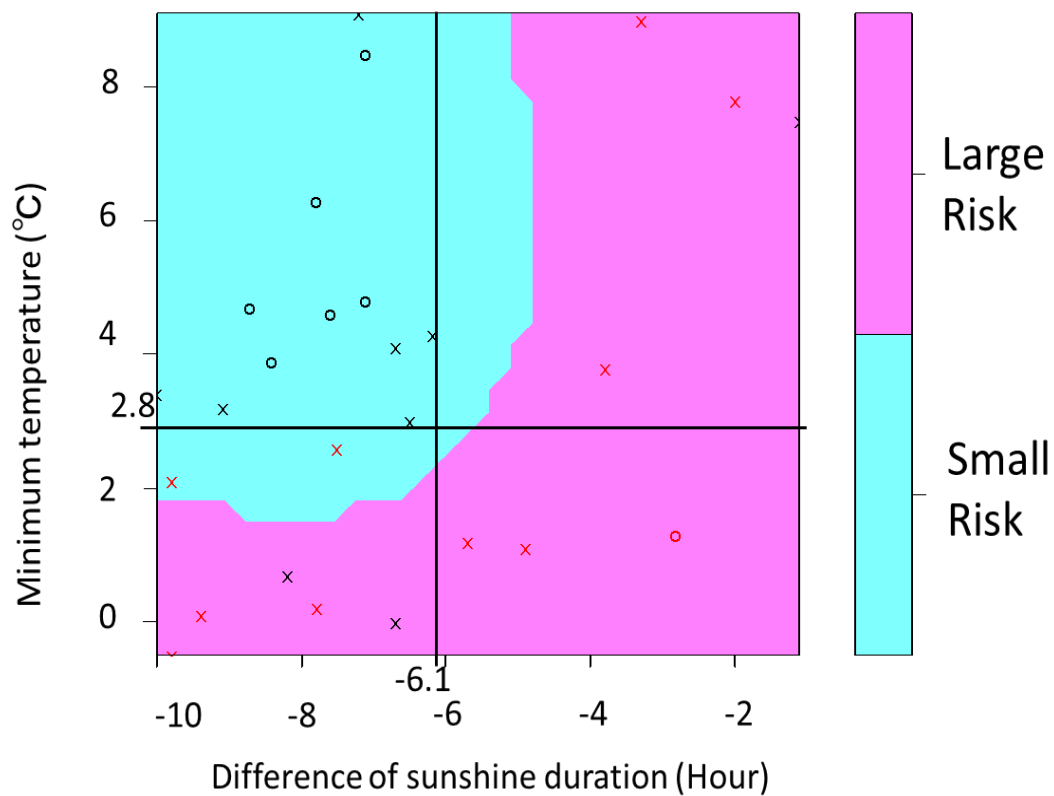


図 1 0 雨天における SVM と決定木の比較

○：サポートベクター

×：実データ

(赤：「リスク・大 (Large Risk)」, 黒；「リスク・小 (Large Small)」

赤色のエリア：SVMにより「リスク・大」と判別されたエリア

青色のエリア：SVMにより「リスク・小」と判別されたエリア

実線：決定木の分岐基準

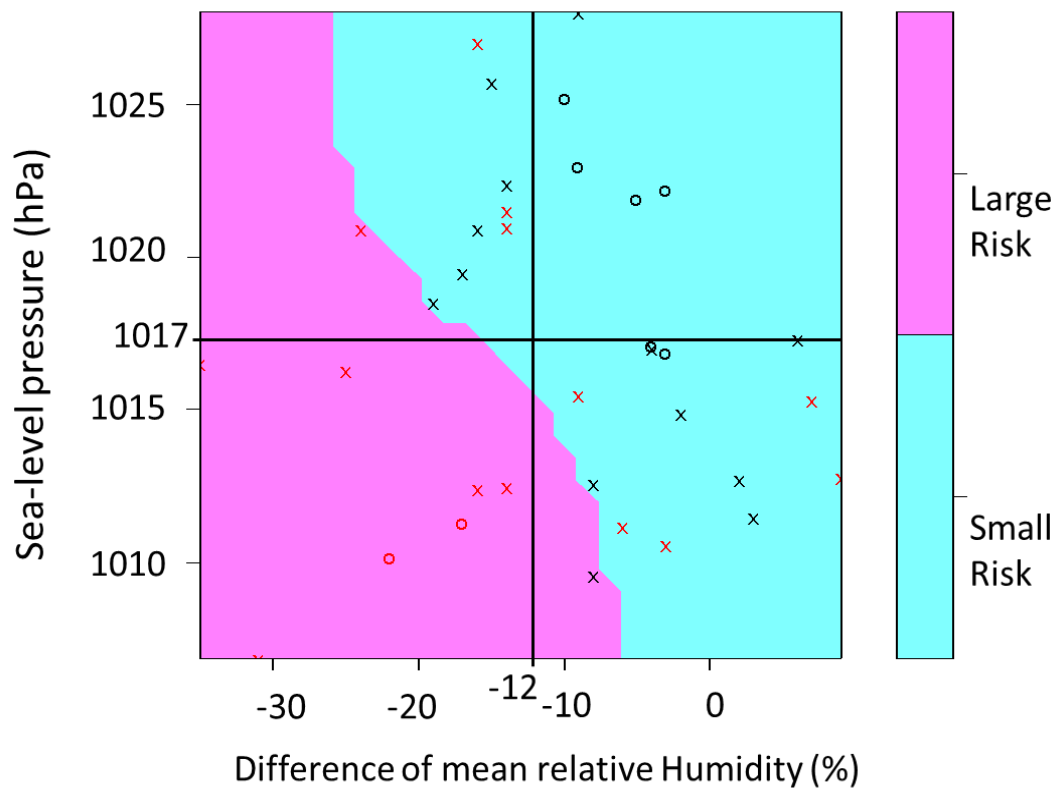


図 1 1 西高東低における SVM と決定木の比較

○：サポートベクター

×：実データ

(赤：「リスク・大 (Large Risk)」, 黒；「リスク・小 (Large Small)」

赤色のエリア：SVMにより「リスク・大」と判別されたエリア

青色のエリア：SVMにより「リスク・小」と判別されたエリア

実線：決定木の分岐基準



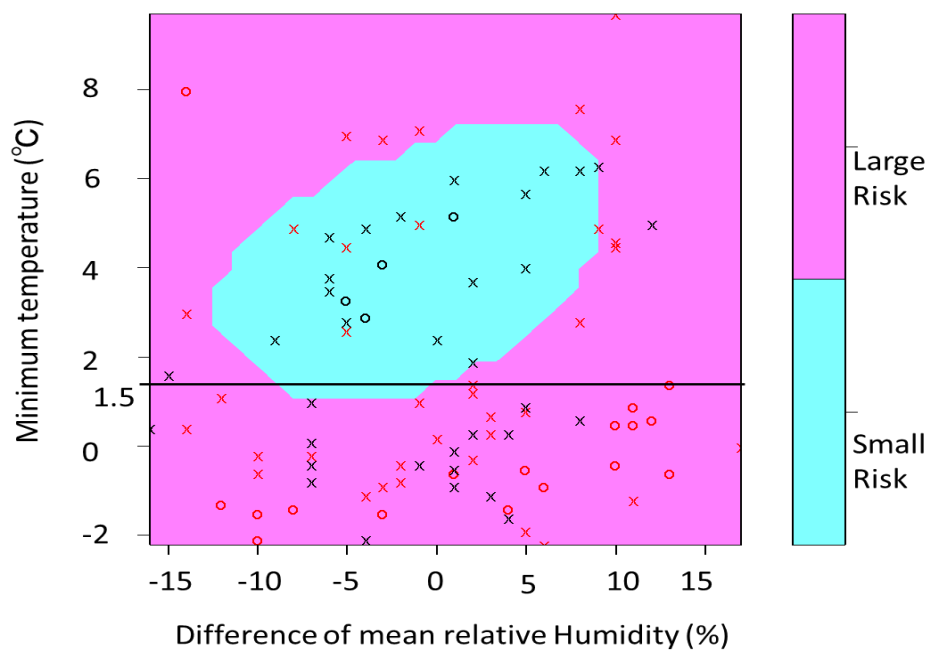


図 1 2 晴天における SVM と決定木の比較

○：サポートベクター

×：実データ

(赤：「リスク・大 (Large Risk)」, 黒；「リスク・小 (Large Small)」)

赤色のエリア：SVMにより「リスク・大」と判別されたエリア

青色のエリア：SVMにより「リスク・小」と判別されたエリア

実線：決定木の分岐基準

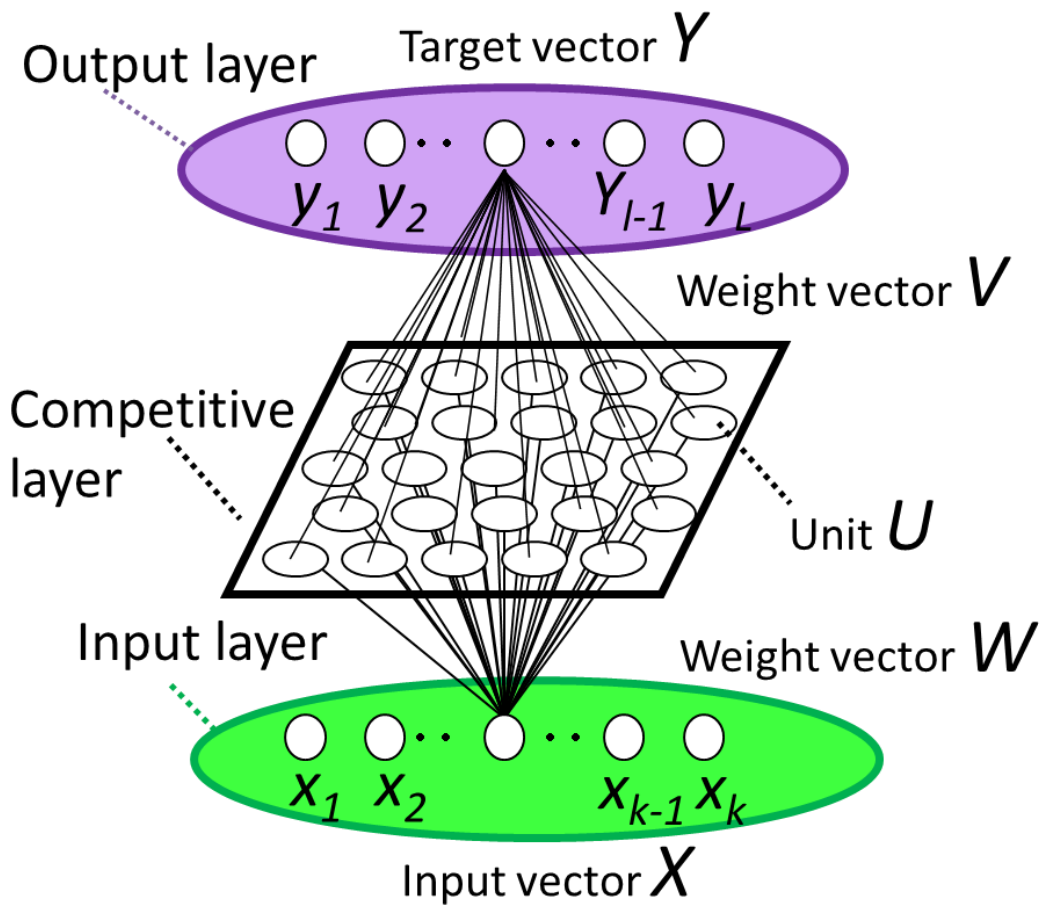


図 1 3 双方向 SOM の概念図

双方向 SOM は, Input Layer (入力層), Competitive Layer (競合層), Output Layer (出力層) で構成されている. 競合層において, 入力層と出力層の重みベクトルが学習されることで, 多次元の重みベクトルが二次元平面上に反映される.

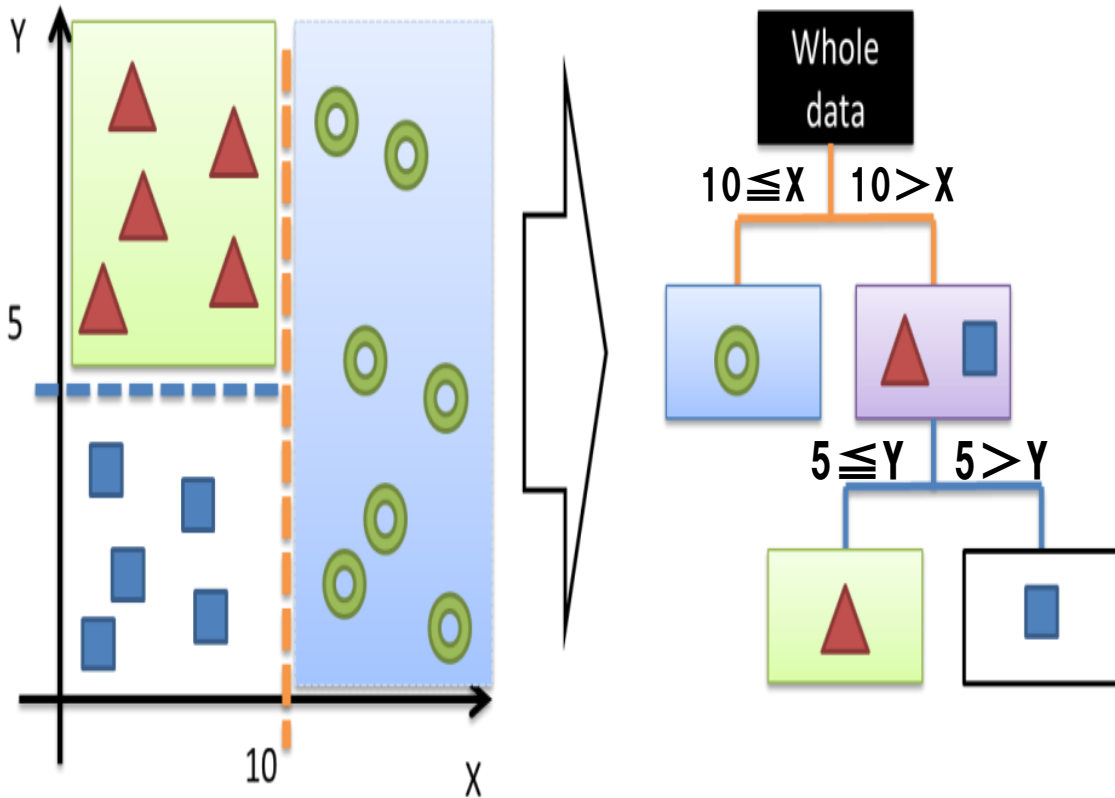


図 14 決定木分類の概念図

全体集合に対して予測変数を用いて、段階的にグループ分けを行い、要素を分類する。全体集合の X 軸 10 を基準として、10 以上を緑の◎と分類している。また、X 軸 10 を基準として 10 未満については、Y 軸 5 を基準として、5 以上を▲とし、5 未満を■と分類することができる。

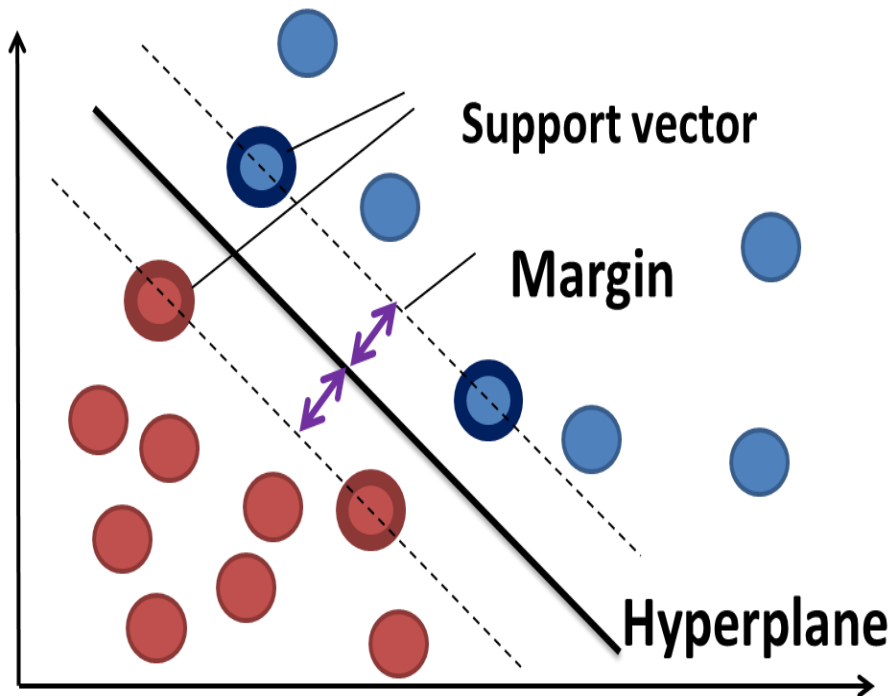


図 15 サポートベクターマシンのマージン最大化

識別境界の両側の破線上の点，サポートベクトル (Support vector) を基準とし，そのユークリッド距離が最も大きくなるような位置に識別境界を設定する．サポートベクトルと識別境界との距離であるマージン (Margin) を最大化するように超空間平面 (Hyper-plane) を構築しクラス分類を行っている．

表 1 脳梗塞の発症に影響を及ぼす可能性がある気象要素を新たに算出した。

気象パラメーター(単位)	=	対象日の気象情報	—	対象日前日の気象情報
平均海面気圧前日差 (hPa)	=	対象日の平均海面気圧	—	対象日前日の平均海面気圧
平均気温前日差 (°C)	=	対象日の平均気温	—	対象日前日の平均気温
最高気温前日差 (°C)	=	対象日の最高気温	—	対象日前日の最高気温
最低気温前日差 (°C)	=	対象日の最低気温	—	対象日前日の最低気温
平均湿度前日差 (%)	=	対象日の平均湿度	—	対象日前日の平均湿度
平均風速前日差 (m/s)	=	対象日の平均風速	—	対象日前日の平均風速
平均雲量前日差 (10段階)	=	対象日の平均雲量	—	対象日前日の平均雲量
日照時間前日差(時間)	=	対象日の日照時間	—	対象日前日の日照時間
平均蒸気圧前日差(hPa)	=	対象日の平均蒸気圧	—	対象日前日の平均蒸気圧

表2 日本気象協会が作成している「天気図型分類表」と、本研究により得られた気候パターンの分類結果との比較

	A	B	C	total
NHSL	2	5	9	16
EHWL	2	0	2	4
SHEL	2	0	4	6
WHEL	12	49	84	145
HP	4	4	53	61
LP	23	2	14	39
total	45	60	166	271

NHSL; North High and South Low pressure distribution type

EHWL; East High and West Low pressure distribution type

SHEL; South High and East Low pressure distribution type

WHEL; West High and East Low pressure distribution type

HP; High atmospheric pressure

LP; Low atmospheric pressure

## Biowetter - Deutschland (Heute): Allgemeines Befinden

In Deutschland, Österreich und der Schweiz sind Millionen Menschen wetterfühlig, d.h. sie reagieren mehr oder weniger stark auf Wetteränderungen. Meist liegen bei diesen Menschen bereits Vorerkrankungen wie Herz-Kreislaufprobleme oder rheumatische Beschwerden vor. Diese können durch bestimmte Wetterlagen beeinflusst werden. Beim Biowetter von wetter.net bekommen Sie interessante Anhaltspunkte über den aktuellen Einfluss der Wetterlage auf Ihr Wohlbefinden vermittelt.



ドイツでは数社が気象情報と併せて医学気象予報を配信している。

例 ; wetter.net (<http://www.wetter.net/biowetter/>) ここでは、25-30 項目の疾患について、4 段階で発症リスクを警告している。

# Relationship between Onset of Cerebral Infarction and Classification of Meteorological Elements

Mitsuhiro Kojima<sup>1</sup>, Norimitsu Ogasawara<sup>2</sup>, Hiroshi Morimoto<sup>1</sup>

<sup>1</sup>Graduate School of Environmental Studies, Nagoya University, Nagoya, Japan

<sup>2</sup>Japan Weather Association, Tokyo, Japan

Email: [kojima.mitsuhiro@e.mbox.nagoya-u.ac.jp](mailto:kojima.mitsuhiro@e.mbox.nagoya-u.ac.jp), [hiroshim@info.human.nagoya-u.ac.jp](mailto:hiroshim@info.human.nagoya-u.ac.jp)

Received 27 February 2014; revised 31 March 2014; accepted 8 April 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

It is well recognized that there are many kinds of diseases that are affected by meteorological factors. This study aims at investigating the relation between the onset of cerebral infarction (CI) and climatic elements. Databases of medical emergency transfers in Nagoya during the winter seasons of 2002-2004 were reviewed. We applied a bi-directional self-organizing map (SOM), cluster analysis and decision tree to estimate the effects of climatic elements on the onset of CI. We classified climatic elements into three patterns, such as “rainy weather”, “west high and east low pressure” and “fair weather” by applying SOM. The risk of CI was estimated more clearly over each pattern. The meteorological variances are associated with the onset of CI. This fact was more accurately observed by classification of climatic patterns based on SOM. This study also suggests that effective warning is possible for the onset of CI, if we classify weather patterns.

## Keywords

Stroke Onset, Weather Chart, Bi-Directional Self-Organizing Map, Decision Tree, Cerebral Infarction Onset

---

## 1. Introduction

The daily evolution of meteorological events influences biological processes and induces human disease. A stroke is considered as one of typical disease affected by climate events [1]. In Germany, health-weather forecasting has been carried by daily newspaper [2]. This study aims to develop health-weather forecasting systems



in Japan. Stroke causes an estimated 5.8 million deaths each year and it is the leading cause of death in Japan.

Many investigators have found that stroke onset peaks happen during the winter [3]-[5]. Weather conditions relevant to the onset of cerebral infarction (CI) have been studied for many years [6]-[9]. Stroke is a generic term for cerebrovascular disorders of sudden onset that comprise CI, cerebral hemorrhage and subarachnoid hemorrhage. Stroke is considered a meteoropathy that worsens when weather conditions change. Relationships between stroke onset and climatic elements have been studied [10]-[12]. The onset of CI is related to climatic elements such as sea-level pressure, mean temperature and sunshine duration. However, quantitative analyses of the risk of CI are challenging.

The present study examined daily transfers of patients with CI resulting from calls to public ambulance emergency services in Nagoya during the winters of 2002-2004 to quantify the effects of weather elements on CI.

We discuss the weather chart type not only climatic elements, and succeeded in understanding the characteristics of the climatic elements from the point of view of meteorology. We used advanced statistical techniques (bi-directional SOM, cluster analysis and decision trees), included several climatic elements in our models and examined the incidence of stroke.

Several studies have reported the relation between stroke and climate elements using logistic regression [10] or multivariate model [13]. The present paper tried to take classification of weather patterns in the method of decision tree.

## 2. Subjects and Methods

### 2.1. Area

This study proceeded in Nagoya (latitude 35°11'N, longitude 136°55'E), a typical commercial city of about 2.2 million inhabitants located in a Pacific climatic zone. Summers are hot and humid, whereas winters are cold, mostly clear and dry.

### 2.2. Database

We obtained information about the numbers of persons transferred to hospital with cerebral infarction (CI) from the database of the Nagoya public ambulance emergency service during December, January and February of 2002-2004.

### 2.3. Meteorological Data

Daily meteorological data were obtained from the Japan Meteorological Agency. Principal component analysis (PCA) was applied to climatic elements to detect correlations. Since the PCA determined that differences in solar radiation and the duration of sunshine closely correlated with other elements, they were removed to avoid multicollinearity between winter climatic elements. PCA was performed with statistical software SPSS 16, using VARIMAX rotation.

We collected data about the following elements: sea-level pressure, mean, minimum and maximum temperatures, diurnal temperature range (maximum-minimum temperature), mean vapor pressure, mean relative humidity, mean wind speed, sunshine duration, solar radiation and precipitation. We then determined differences in the following elements: sea-level pressure, mean, minimum and maximum temperatures, vapor pressure, mean relative humidity, mean wind speed, cloud cover and sunshine duration.

### 2.4. Weather Chart Type

We used the “weather chart classification” (weather chart) created by the Japan Weather Association. These charts classify climatic conditions based on the position and distance of weather fronts from Nagoya, such as the type of high/low pressure distribution and nature of atmospheric pressure, as follows: east high and west low (EHWL); north high and south low (NHSL); south high and east low (SHEL); west high and east low (WHEL) along with high (HP) and low (LP) atmospheric pressure.

Low pressure type is determined by the positional relationship between the pressure gradient. Type of front is determined by front type, direction, and the distance to the front.

### 3. Statistical Methods

#### Learning Algorithm of Bi-Directional SOM

The SOM is used to make prediction and trend analysis by visualizing complex high-dimensional data onto simple two-dimensional grids [14]. SOM is comprised of an input layer and a competitive layer.

Bi-directional SOM uses a supervised learning algorithm achieved by adding one more layer (called output layer) to the normal SOM. Thus, bi-directional SOM consists of three layers, input, competitive and output layers (Figure 1). Input layer has normally high-dimensional input vectors ( $x_k \in R^K$ ). Output layer has a target vector ( $y_n \in R^n$ ) acting as a teacher signal.

In the bi-directional SOM, we use the Euclidean distance in order to measure the similarity of the weight vectors. The pair of weight vectors ( $w_c(t), v_c(t)$ ) of the winner unit ( $u_c$ ) satisfies the following Equation (1)

$$\|x(t) - w_c(t)\| + \|y(t) - v_c(t)\| = \min_i \{ \|x(t) - w_i(t)\| + \|y(t) - v_i(t)\| \}, \quad (1)$$

The weight vectors of units in the neighborhood of the winner unit are also updated according to the distance from the winner unit.

Weight vector ( $w_i(t+1), v_i(t+1)$ ) at time ( $t+1$ ) of the unit ( $u_i$ ) is defined as follows, using a weight vector ( $w_i(t), v_i(t)$ ) at time ( $t$ ).

$$\begin{aligned} w_i(t+1) &= w_i(t) + h_{ci}(t) \{ x(t) - w_i(t) \} \quad u_i \in N_c(t), \\ v_i(t+1) &= v_i(t) + h_{ci}(t) \{ y(t) - v_i(t) \} \quad u_i \in N_c(t), \end{aligned} \quad (2)$$

where

$$h_{ci}(t) = \alpha(t) \exp \left( - \frac{\|r_c(t) - r_i\|^2}{2\sigma^2(t)} \right).$$

$N_c(t)$ : Neighboring of the winner unit ( $u_c$ );

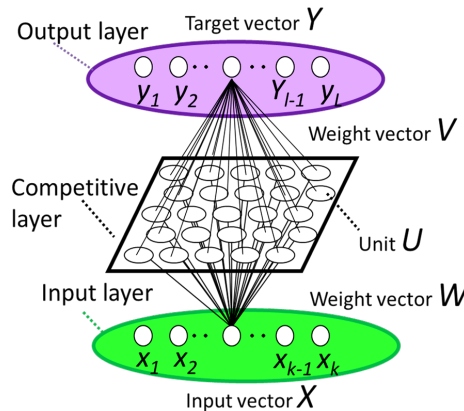
$\sigma(t)$ : Positive constant;

$r_c(t) - r_i$ : Position vector of the two-dimensional plane.

Thus, the feature of the input and teacher vector is reflected in each weight vector (2). Thus, two kinds of weight vectors arranged in each unit in two-dimensional plane become the final results of bi-directional SOM.

In this study, we used a bi-directional SOM to classify the climate pattern with respect to the risk of CI. And then, we used bi-directional SOM with 20-dimensional climatic elements as input vector, the transferred number of CI as target vector.

In this study, SOM is performed by using the statistical software “R” and its functions “som” and “xyf” in the package of “kohonen”. The function “som” calculates a basic SOM by using only input data, and the function



**Figure 1.** Conceptual diagram of bi-directional SOM. Bi-directional SOM consists of three layers, input, competitive and output layers.

“xyf” calculates bi-directional SOM by adding target data. These functions use Gaussian by default.

Input and output data is normalized in advance. The grids of units are set to  $5 \times 5$  lattices. The number of iterations is set to 1000.

We then performed a cluster analysis of the climatic elements based on the results of the bi-directional SOM and classified them into climate patterns A, B and C.

In this study, we used Ward’s method in cluster analysis. As a result of cluster analysis, the feature of climate patterns was classified into three clusters.

## 4. Result

### 4.1. Climate Pattern

We see that the bi-directional SOM can contribute to trend-analysis and classification of relationship between complex 20-dimensional climatic elements and transferred number of CI onto simple two-dimensional grid.

We compared our result and that by Japan Weather Association. Thus, we could identify our classification (by SOM) with the weather patterns by Japan Weather Association.

Climate pattern A essentially comprised rainy weather with high levels of precipitation, a short duration of sunshine and low levels of solar radiation. Climate pattern B was classified as having a west high and east low pressure distribution (WHEL). Climate pattern C (fair weather) was classified as comprising high atmospheric pressure (Table 1).

### 4.2. Estimation of the Risk of CI

One “unit” is a group classified by the SOM. Small (low risk of CI) and large (high risk of CI) units contain input vectors with daily translation numbers  $<6$  and  $>7$ . Average and standard deviation of transport number of CI was 5.3 and 2.4 respectively. From these results, we determined daily transport numbers of CI more than 7 as high risk.

### 4.3. Risk of CI in Climate Pattern A (“Rainy Weather”)

We analyzed CI under conditions of climate pattern A (rainy weather) using a decision tree to determine differences in low and high risk. Minimal temperature in the first node branched at  $3.1^{\circ}\text{C}$ .

Risk was reduced and increased when the minimum temperature was  $>3.1^{\circ}\text{C}$  and  $<3.1^{\circ}\text{C}$  respectively.

### 4.4. Risk of CI in Climate Pattern B (“WHEL”)

We analyzed individuals with CI transferred under a WHEL by the similar way as the pattern A. Difference of mean relative humidity in the first node branched at  $-21.5\%$ . Risk was reduced and increased when difference of mean relative humidity was  $>-21.5\%$  and  $<-21.5\%$  respectively.

### 4.5. Risk of CI in Climate Pattern C (“Fair Weather”)

We analyzed the data of this pattern similarly. Diurnal temperature range in the first node branched at  $8.55^{\circ}\text{C}$ .

**Table 1.** Comparison of weather chart type and climate patterns.

Weather chart type	Climate patterns		
	A	B	C
NHSL	2	5	16
EHWL	2		4
SHEL	2		6
WHEL	12	49	145
HP	4	4	61
LP	23	2	39

Risk was respectively reduced and increased at diurnal temperatures above and below 8.55°C.

#### 4.6. Predicted Risk of CI Onset

The present study compared the predictive accuracy of decision trees, bi-directional SOM, and cluster analysis with that of decision tree alone. The predictive accuracy of decision tree alone (without weather classifications) was 77.5%. Adding our climate classification improved this to 95.6%, 81.3% and 95% during rainy weather, fair weather, and during a west high and east low pressure distribution, respectively.

### 5. Discussion

Stroke is considered to be a meteorotropism. Associations between the onset of CI and the climatic elements, such as air pressure, temperature and humidity have been investigated [15] [16].

However, these reports did not sufficiently consider pressure patterns and the position of weather fronts. We classified climate patterns and elements, and then estimated the risk of CI using bi-directional SOM and multi-dimensional data. Bi-directional SOM allowed the classification of weather elements according to the risk of CI as “rainy weather”, “west high and east low pressure distribution” and “fair weather”. We found that bi-directional SOM is effective for understanding complicated changes in climatic elements. In addition, the application of decision tree allowed a more precise estimation of the risk of CI associated with each climatic condition.

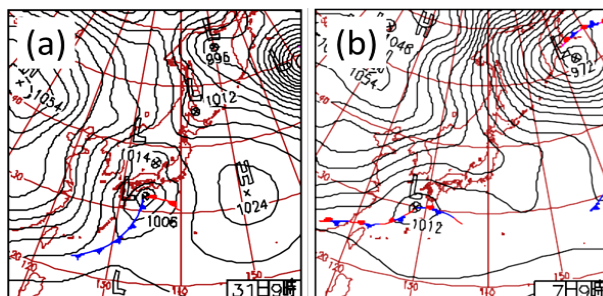
The risk of CI was higher in rainy weather, for example, when a cold front passed through Nagoya (**Figure 2**). Precipitation and low temperature also increased risk. On the other hand, risk was low when a front was stationary. We can say that the risk during rainy weather is affected by temperature and front lines.

Here are examples of CI risk under climate pattern B. The risk became increased immediately after a front passed through Nagoya, that is, on the day after precipitation (**Figure 3**). On the other hand, risk decreased when a front passed through followed by a distant high pressure front.

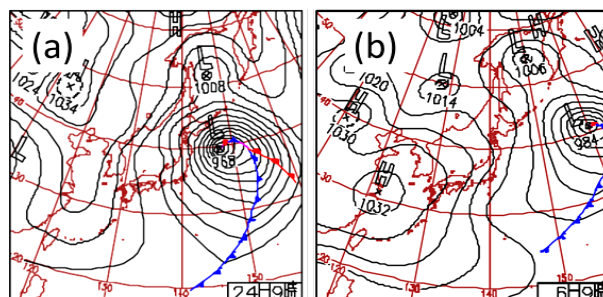
We found that over 10 individuals were transported during fair weather and that the risk of CI increased at higher temperatures, or when the diurnal temperature range was 11.7°C. Heat stress might be caused by transitions in temperatures. High/low temperatures might cause psychological stress, which in turn would increase blood pressure and thus the risk of CI. Here are examples of increased CI risk under climate pattern C (**Figure 4**). Heat stress appeared to affect the autonomic nervous system. Rapid changes in temperature seemed to exert considerable physical stress. Risk of CI was high when temperatures increased, or when the diurnal temperature range was 11.7°C.

Judging from these results, the mechanism of the onset of CI varies with climate patterns. We found that the risk of CI onset can be estimated more effectively from decision trees which are applied after classifying climate patterns in bi-directional SOM, than by using conventional decision trees that only consider weather conditions without any classification of climatic elements. We classified evolutionary patterns of meteorological events into three patterns. This enabled to analyse more clearly the relation of the onset of CI and climate elements. The method of decision tree was improved by applying it over each classified climate pattern.

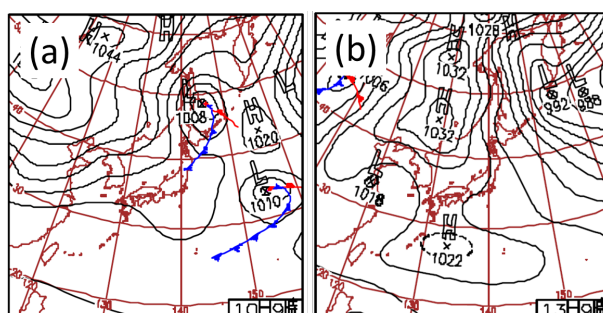
We can conclude that classification of climate patterns improves the accuracy of health-weather forecasting.



**Figure 2.** Typical example in Group A, “rainy weather”. (a) Risk of cerebral infarction becomes higher. On December 31, 2004, twelve persons were reported to have been transported; (b) Risk of cerebral infarction becomes lower. On December 7, 2002, one person was reported to have been transported.



**Figure 3.** Typical example in Group B, “West High and East Low pressure distribution type (WHEL)”. (a) Risk of cerebral infarction becomes higher. On December 24, 2003, eleven persons were reported to have been transported; (b) Risk of cerebral infarction becomes lower. On February 6, 2003, two persons were reported to have been transported.



**Figure 4.** Typical example in Group C, “fair weather”. (a) Risk of cerebral infarction becomes higher. On February 10, 2003, thirteen persons were reported to have been transported; (b) Risk of cerebral infarction becomes lower. On January 13, 2002, one person was reported to have been transported.

Thus effective warnings can be possible based on our method.

## Acknowledgements

We thank the public ambulance emergency service of Nagoya for providing the patients’ data.

## References

- [1] Lanska, D.J. and Hoffmann, R.G. (1999) Seasonal Variation in Stroke Mortality Rates. *Neurology*, **52**, 984. <http://dx.doi.org/10.1212/WNL.52.5.984>
- [2] Goggins, W.B., Woo, J., Ho, S., Chan, E.Y. and Chau, P.H. (2012) Weather, Season, and Daily Stroke Admissions in Hong Kong. *International journal of biometeorology*, **56**, 865-872. <http://dx.doi.org/10.1007/s00484-011-0491-9>
- [3] Walach, H., Schweickhardt, A. and Bucher, K. (2002) Does Weather Modify Headaches? An Empirical Evaluation of Bio-Weather Categorization. *Schmerz (Berlin, Germany)*, **16**, 1-8. <http://dx.doi.org/10.1007/s004820100066>
- [4] Inagawa, T. (2002) Seasonal Variation in the Incidence of Aneurysmal Subarachnoid Hemorrhage in Hospital-and Community-Based Studies. *Journal of neurosurgery*, **96**, 497-509. <http://dx.doi.org/10.3171/jns.2002.96.3.0497>
- [5] Fischer, T., Johnsen, S.O.R.P., Pedersen, L., Gaist, D., Sørensen, H.T. and Rothman, K.J. (2004) Seasonal Variation in Hospitalization and Case Fatality of Subarachnoid Hemorrhage—A Nationwide Danish Study on 9367 Patients. *Neuroepidemiology*, **24**, 32-37. <http://dx.doi.org/10.1159/000081047>
- [6] Chen, R., Wang, C., Meng, X., Chen, H., Thach, T.Q., Wong, C.M. and Kan, H. (2013) Both Low and High Temperature May Increase the Risk of Stroke Mortality. *Neurology*, **81**, 1064-1070. <http://dx.doi.org/10.1212/WNL.0b013e3182a4a43c>
- [7] Kawai, K., Nonaka, K., Suzuki, H., Kirino, T. and Tamura, A. (2001) Differential Effects of Activity and Climate on

Onset of Subarachnoid Hemorrhage. *Neurologia Medico-Chirurgica*, **41**, 229-236.  
<http://dx.doi.org/10.2176/nmc.41.229>

- [8] Wang, H., Matsumura, M., Kakehashi, M. and Eboshida, A (2006) Effects of Atmospheric Temperature and Pressure on the Occurrence of Acute Myocardial Infarction in Hiroshima City, Japan. *Hiroshima Journal of Medical Sciences*, **55**, 45-51.
- [9] Lim, Y.H., Kim, H. and Hong, Y.C. (2013) Variation in Mortality of Ischemic and Hemorrhagic Strokes in Relation to High Temperature. *International journal of Biometeorology*, **57**, 145-153.  
<http://dx.doi.org/10.1007/s00484-012-0542-x>
- [10] Setzer, M., Beck, J., Hermann, E., Raabe, A., Seifert, V., Vatter, H. and Marquardt, G. (2007) The Influence of Barometric Pressure Changes and Standard Meteorological Variables on the Occurrence and Clinical Features of Subarachnoid Hemorrhage. *Surgical Neurology*, **67**, 264-272. <http://dx.doi.org/10.1016/j.surneu.2006.06.060>
- [11] Hong, Y.C., Rha, J.H., Lee, J.T., Ha, E.H., Kwon, H.J. and Kim, H.O. (2003) Ischemic Stroke Associated with Decrease in Temperature. *Epidemiology*, **14**, 473-478. <http://dx.doi.org/10.1097/01.ede.0000078420.82023.e3>
- [12] Woo, J., Kay, R. and Nicholls, M.G. (1991) Environmental Temperature and Stroke in a Subtropical Climate. *Neuroepidemiology*, **10**, 260-265. <http://dx.doi.org/10.1159/000110282>
- [13] Lejeune, J.P., Vinchon, M., Amouyel, P., Escartin, T., Escartin, D. and Christiaens, J.L. (1994) Association of Occurrence of Aneurysmal Bleeding with Meteorologic Variations in the North of France. *Stroke; a Journal of Cerebral Circulation*, **25**, 338-341. <http://dx.doi.org/10.1161/01.STR.25.2.338>
- [14] Kohonen, T. (1982) Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, **43**, 59-69. <http://dx.doi.org/10.1007/BF00337288>
- [15] Oberg, A.L., Ferguson, J.A., McIntyre, L.M. and Horner, R.D. (2000) Incidence of Stroke and Season of the Year: Evidence of an Association. *American journal of epidemiology*, **152**, 558-564. <http://dx.doi.org/10.1093/aje/152.6.558>
- [16] Liang, W., Hu, Y., Kasabov, N. and Feigin, V. (2011) Exploring Associations between Changes in Ambient Temperature and Stroke Occurrence: Comparative Analysis Using Global and Personalised Modelling Approaches. In *Neural Information Processing*, Springer, Berlin, Heidelberg, 129-137.

1 ***Title page :***

2 The prediction of the onset of risk of cerebral infarction by climatic elements using  
3 Support vector machine

4  
5  
6 Mitsuhiro Kojima, Hiroshi Morimoto, Tamihito Nishida

7  
8  
9  
10 Graduate School of Environmental Studies, Nagoya University  
11 Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan

12  
13 Address correspondence to Mitsuhiro Kojima  
14 Student, Graduate School of Environmental Studies, Nagoya University,  
15 Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan  
16 E-mail: [kojima.mitsuhiro@e.mbox.nagoya-u.ac.jp](mailto:kojima.mitsuhiro@e.mbox.nagoya-u.ac.jp)  
17 E-mail: [hiroshim@info.human.nagoya-u.ac.jp](mailto:hiroshim@info.human.nagoya-u.ac.jp)

18  
19  
20  
21  
22  
23  
24  
25  
26

***Keywords:***

onset of cerebral infarction • support vector machine (SVM) • climatic elements •  
bi-directional Self-Organizing maps • onset of stroke

27 ***Background and Purpose***

28 Several studies have reported the effects of season or climatic elements on the onset  
29 of stroke.

30 The existence of relationship between the onset of cerebral infarction and  
31 temperature has been pointed by a number of studies, and the difference of seasonal  
32 onset has been considered.

33 The present study was undertaken to develop a high accuracy method to predict the  
34 onset of cerebral infarction by support vector machine.

35

36 ***Methods***

37 This study uses the data of climatic elements and the numbers of persons transferred  
38 to hospital with cerebral infarction in the winter of 2002-2004, Nagoya City. We  
39 classified the climate pattern by bi-directional self-organizing maps (SOM). In each  
40 climate pattern, we compared the prediction accuracy of the method by Support Vector  
41 Machine (SVM) and the method by decision tree.

42

43 ***Results***

44 The climate patterns were classified into three types (fine weather: n=45, "n" is  
45 Number of days in the climate pattern, "high pressure in the west and low pressure in  
46 the east": n=60, rainy weather: n=166) by using the method of bi-directional SOMs. In  
47 each climate pattern, the prediction accuracy of decision tree and SVM were  
48 compared, and we found prediction accuracy of the onset of cerebral infarction was  
49 improved by using SVM.

50

51 ***Conclusions***

52 This paper proposed a new method by SVM to predict the onset of cerebral infarction  
53 with high accuracy.

54

55



56 ***Introduction***

57

58 There are many studies on the effects of seasonal or climatic elements on the onset of  
59 stroke (Lanska et al.1999; Chen et al. 2013; Kawai et al. 2001; Lim et al. 2013; Oberg  
60 et al. 2000). The researchers in Germany, studied the relationship between climatic  
61 elements and onset of stroke to predict of onset of stroke in the 1950s, and they  
62 developed first “Bio-weather forecast (health-weather forecasting systems)” in the  
63 world (Walach et al. 2002; Lim et al. 2013).

64 In Japan, Hiroshima prefectural medical association studied the relationship between  
65 onset of stroke and climatic elements using the database of public ambulance  
66 emergency service of Hiroshima city. In Hiroshima, the bio-weather forecast of stroke  
67 is announced every day (Wang et al. 2006).

68 In this paper, relationships between onset of cerebral infarction and climatic elements  
69 were analyzed by SVM. It is known that SVM is great performance as versatile and  
70 classification performance, taking advantage of the characteristics of the non-linear  
71 model and margin maximization,

72 We analyzed quantitatively the onset prediction accuracy of CI. And, in order to  
73 discuss the practicality of statistical methods, we compared with prediction accuracy  
74 of decision tree.

75 Furthermore, the classification in climate patterns from some conditions of climatic  
76 elements by bi-directional SOMs, to determine the prediction accuracy in climate  
77 pattern for each.

78

79

80

81 ***Subjects and Methods***

82

83 ***Subjects***

84

85 Nagoya city is located in a Pacific climatic zone. It is hot and humid in summer,  
86 whereas in winter it is cold, mostly clear and dry.

87 We used the database of the numbers of persons transferred to hospital by ambulance  
88 during 2002-2004. Among them, we selected data of persons who were diagnosed as  
89 cerebral infarction (CI) during winter seasons (December, January and February).

90 Each report of conveyance includes the data of “name of disease”, “location”,  
91 “age”, “date”, “dispatch time”, “degree injuries”, “gender” and “injury part”.

92 The number of patients who were diagnosed as CI was 2,875 (male: 1,756, female:  
93 1119). Among them, in January, the number of patients is 1,028 (male: 633, female:  
94 395). In February and December, these numbers were 902(male: 633, female: 395),  
95 945 (male: 556, female: 389), respectively.

96

97 ***Meteorological data***

98

99 Daily meteorological data were downloaded from the database of Japan  
100 Meteorological Agency.

101 Principal component analysis (PCA) was applied to avoid multicollinearity between  
102 winter climatic elements. PCA was performed with statistical software SPSS 16.

103 In this study, we used these 20 climatic elements (Table 1.).

104

105

106 ***Statistical Analysis***

107 ***Self-organizing map***

108 The SOM (self-organizing map) has been applied in various fields including  
109 meteorology, medicine and engineering. The SOM was used to make prediction and  
110 trend analysis by visualizing complex high-dimensional data onto simple  
111 two-dimensional grids (Kohonen et al. 1982; Kohonen et al. 1990). Bi-directional  
112 SOM uses a supervised learning algorithm achieved by adding one more layer (called  
113 output layer) to the normal SOM. Thus, bi-directional SOM consists of three layers,  
114 input, competitive and output layers. Input layer is the same as that of SOM.

115 In this study, SOM is performed by using the statistical software “R” and its  
116 function “xyf” in the package of “kohonen”. Input and output data are normalized in  
117 advance. The grids of units are set to 5×5 lattices. The number of iterations is set to  
118 1000.

119

120 ***Support Vector Machine***

121 Support Vector Machine (SVM) is an effective classifier for use in classification  
122 into two classes (Figure 1).

123 The aim of SVM is to find a hyperplane (or more generally hypersurface) which  
124 divides the given two classes (such as the class of high risk and the class of low risk)  
125 in the training data. The minimum of distances between the hyperplane and points of  
126 each class is called the "margin". The points which give the minimum distance are  
127 called "support vectors". The support vectors are thus regarded as the boundary of  
128 each class. The best hyperplane is that of the largest margin, and the process to find the  
129 best hyperplane is called "maximizing margin" (Cristianini et al. 2000; Tripathi et al.  
130 2006).

131 In this study, SVM was performed by using the typical “RBF kernel trick function”.  
132 SVM is performed by using the statistical software “R” and its functions “e1071” in  
133 the package of “svm”.

134

135 ***Prediction accuracy***

136 In this section, we verified quantitatively the effectiveness and versatility of our  
137 method by SVM. To discuss the practicality of a statistical technique, we determine  
138 the prediction accuracy.

139 First, we classified the data according to the number of persons transported to  
140 hospitals (Figure 2). If the daily transport number of CI is more than 6, then such  
141 case is called the case of “Large Risk”. If the daily transport number of CI is less than  
142 3, then such case is called the case of “Small Risk”. And all the other cases are called  
143 the cases of “Middle Risk”. In this study, we used only “Large Risk” and “Small Risk”,  
144 to predict of the high and low risk of CI. This is because CI is caused by many other  
145 reasons rather than weather; the cases of “Middle Risk” may not be influenced by  
146 weather and such cases can be considered as "noise" for our study.

147 The objective variable of SVM and decision tree is set for the onset risk of CI  
148 (“Large Risk”, “Small Risk”) and the dependent variables are set for climatic elements.  
149 In decision tree, we divided the data into two groups, learning data and test data.  
150 Learning data were extracted at random 70% from the data, and other of 30% was  
151 selected for test data.

152 In this study, the branch of decision tree was planted up to the second node to  
153 enhance versatility. We constructed a model by the algorithm by a decision tree  
154 applied to the learning data, and then the model is applied to the test data for the test  
155 of the model. This process was repeated three times. The average of the correct rates  
156 of these 3 times repeats as "the prediction accuracy" of decision tree.

157 We now proceed to apply SVM. The dependent variables of SVM were selected to be  
158 the climatic elements which appeared as a branch of decision tree in the first and  
159 second node. Prediction accuracy of SVM was calculated by k-fold cross-validation  
160 method (k=3).

161 Finally, we do the same process restricting the data over the specific climatic patterns,  
162 which are determined by bi-directional SOM.

163

164

165

166 **Results**

167  
168 ***Classification of climate patterns by bi-directional SOM***

169  
170 First, we classify climate patterns by SOM taking into consideration the effect of CI.  
171 For this purpose, a bi-directional SOM was better than normal SOM. The data of 20  
172 climatic elements, (such as the maximum temperature, minimum temperature and  
173 sea-level pressure etc.) during the terms considered. In applying the bi-directional  
174 SOM, we set 20 climatic elements to be the first layer of SOM and set the number of  
175 persons conveyed by ambulance to be the third layer. The second layer is the so-called  
176 "competitive layer". Second, we apply cluster analysis to classify the results of SOM.

177 Finally we get three climate patterns, "rainy weather" (pattern A), "high pressure in  
178 the west and low pressure in the east" (pattern B) and "fair weather" (pattern C).

179 In "rainy weather", low pressure is approaching and precipitation was observed. The  
180 pattern "high pressure in the west and low pressure in the east" appears mostly in  
181 winter in Nagoya. In "fine weather", high pressure is close (Figure 3.).

182 The "prediction accuracy" of the models is defined as follows. We apply the test data  
183 to the model (SVM or Decision tree). Then the model gives the result of prediction,  
184 "Large Risk" or "Small Risk". If the prediction is the same as the observed data, then  
185 the prediction is said to be correct. Gathering all of the correct prediction, the  
186 "prediction accuracy" is defined to be the proportion of these correct predictions in all  
187 of the test data.

188 First we compare the prediction accuracy of SVM and decision tree without the  
189 classification of climate patterns. The prediction accuracy was 53% for the decision  
190 tree and 56% for SVM. Thus the prediction accuracy of SVM is better than the  
191 decision tree, but both are not accurate enough (Figure 4.).

192 Second, we compare the prediction accuracy using the classification of climate  
193 patters. In "rainy weather", the prediction accuracy of decision tree was 62.5% and  
194 that of SVM was 69.2%. The results of classification by SVM and decision tree  
195 resulted in almost the same, but strictly speaking, the prediction of SVM accuracy was

196 better (Figure 5-1). When the difference of sunshine duration (with that of one day  
197 before) was shorter than 5 hours, and the minimum temperature was higher than 2°C,  
198 both SVM and decision tree predicted as "Small Risk". However, when the minimum  
199 temperature was lower than 2°C, the risk was predicted as "Large Risk". As it has been  
200 reported in many papers that there are relationship of stroke and minimum temperature,  
201 it was suggested that there are relationship between the minimum temperature and CI  
202 in our study (Liang et al. 2011; Woo et al. 1991).

203 In "high pressure in the west and low pressure in the east ", the classification criteria  
204 by decision tree tells us that sea-level pressure and difference of mean relative  
205 humidity are main climate elements to determine the risk of CI. Prediction accuracy  
206 by decision tree was 54.5%. Prediction accuracy by SVM was 62.9% (for the same  
207 two variables as above). In this case, the onset of CI was effectively predicted by SVM  
208 rather than decision tree (Figure 5-2). From the results of SVM, it is suggested that the  
209 onset risk increased when the sea-level pressure was reduced. It has not been reported  
210 that relationship between stroke and air pressure, but we could suggest that the onset  
211 risk of CI increases when the air pressure decreases (Lejeune et al. 1994; Setzer et al.  
212 2007).

213 In "fair weather" cases, the prediction accuracy of decision tree was 58.6% and that  
214 of SVM was 67%. The prediction accuracy by SVM was higher than that by decision  
215 tree (Figure 5-3). If the minimum temperature is all most the equal to average level  
216 and the humidity keeps constant, then we can see from the figure that the risk of CI  
217 becomes small. If the minimum temperature is less than or equal to 2°C, the risk of CI  
218 becomes higher. This means that the minimum temperature plays an important role. If  
219 the difference of humidity between two days is higher than  $\pm 10\%$ , then the risk of CI  
220 becomes higher (Hong et al. 2003). This implies that the humidity also plays an  
221 important role for the prediction and that not only the climate conditions of the day in  
222 question but also those of previous day are connected to the onset of CI.

223

224

225 *Discussion*

226

227 First, if we do not use climate patterns, then we saw that both methods SVM and  
228 decision tree were not effective at all. The cases “Large Risk” and “Small Risk” were  
229 not well classified and the correct answer rates were not remarkable. This is due to the  
230 variability of the weather in Nagoya City.

231 Second, we took the three climate patterns into consideration, and applied both  
232 methods of SVM and decision tree restricting the data to each climate pattern, such as  
233 "rainy weather", " west high and east low " and "fair weather".

234 The results of accuracy rate showed that SVM scored better than decision tree over  
235 any of climate patterns. A feature of margin maximization of SVM was effective to  
236 classify the onset risk of CI. In particular, SVM with two selected climate elements  
237 were proven to be most effective for the prediction of CI. Especially for the cases of  
238 the climate pattern "fine weather", the method of decision tree could not predict the  
239 risk of CI.

240 On the contrary, SVM marked far better score than the decision tree and was found to  
241 be effective tool for the prediction of CI. The prediction became possible with the  
242 accuracy of 62.9-69.2% by SVM.

243 In future, more effective choice of climate elements will be discussed in order to  
244 raise the prediction accuracy of the onset of CI.

245

246

247 **References**

248

249 Chen, R., Wang, C., Meng, X., Chen, H., Thach, T. Q., Wong, C. M., & Kan, H.  
250 (2013). Both low and high temperature may increase the risk of stroke  
251 mortality. *Neurology*, 81(12), 1064-1070.

252

253 Cristianini, N., & Shawe-Taylor, J. (2000). An introduction to support vector machines  
254 and other kernel-based learning methods. Cambridge university press.

255

256 Hong, Y. C., Rha, J. H., Lee, J. T., Ha, E. H., Kwon, H. J., & Kim, H. O. (2003).  
257 Ischemic stroke associated with decrease in temperature. *Epidemiology*, 14(4),  
258 473-478.

259

260 Kawai, K., Nonaka, K., Suzuki, H., Kirino, T., & Tamura, A. (2001). Differential  
261 effects of activity and climate on onset of subarachnoid hemorrhage. *Neurologia*  
262 *medico-chirurgica*, 41(5), 229-36.

263

264 Kohonen T. Self-organized formation of topologically correct feature maps. *Biological*  
265 *cybernetics*, 1982, 43.1: 59-69.

266

267 Kohonen T. The self-organizing map. *Proceedings of the IEEE*, 1990, 78.9:  
268 1464-1480.

269 Lanska, D. J., & Hoffmann, R. G. (1999). Seasonal variation in stroke mortality  
270 rates. *Neurology*, 52(5), 984-984.

271

272 Lejeune, J. P., Vinchon, M., Amouyel, P., Escartin, T., Escartin, D., & Christiaens, J. L.  
273 (1994). Association of occurrence of aneurysmal bleeding with meteorologic  
274 variations in the north of France. *Stroke; a journal of cerebral circulation*, 25(2),  
275 338-341.

276

277 Liang, W., Hu, Y., Kasabov, N., & Feigin, V. (2011, January). Exploring associations  
278 between changes in ambient temperature and stroke occurrence: comparative analysis  
279 using global and personalized modelling approaches. In *Neural Information*  
280 *Processing* (pp. 129-137). Springer Berlin Heidelberg.



281

282 Lim, Y. H., Kim, H., & Hong, Y. C. (2013). Variation in mortality of ischemic and  
283 hemorrhagic strokes in relation to high temperature. *International journal of*  
284 *biometeorology*, 57(1), 145-153.

285

286 Oberg, A. L., Ferguson, J. A., McIntyre, L. M., & Horner, R. D. (2000). Incidence of  
287 stroke and season of the year: evidence of an association. *American journal of*  
288 *epidemiology*, 152(6), 558-564.

289

290 Setzer, M., Beck, J., Hermann, E., Raabe, A., Seifert, V., Vatter, H., & Marquardt, G.  
291 (2007). The influence of barometric pressure changes and standard meteorological  
292 variables on the occurrence and clinical features of subarachnoid  
293 hemorrhage. *Surgical neurology*, 67(3), 264-272.

294

295 Tripathi, S., Srinivas, V. V., & Nanjundiah, R. S. (2006). Downscaling of precipitation  
296 for climate change scenarios: a support vector machine approach. *Journal of*  
297 *Hydrology*, 330(3), 621-640.

298

299 Walach, H., Schweickhardt, A., & Bucher, K. (2002). [Does weather modify  
300 headaches? An empirical evaluation of bio-weather categorization]. *Schmerz* (Berlin,  
301 Germany), 16(1), 1-8.

302

303 Wang, H., Matsumura, M., Kakehashi, M., & Eboshida, A (2006). Effects of  
304 atmospheric temperature and pressure on the occurrence of acute myocardial  
305 infarction in Hiroshima City, Japan. *Hiroshima journal of medical sciences*, 55(2),  
306 45-51.

307

308 Woo, J., Kay, R., & Nicholls, M. G. (1991). Environmental temperature and stroke in a  
309 subtropical climate. *Neuroepidemiology*, 10(5-6), 260-265.

310

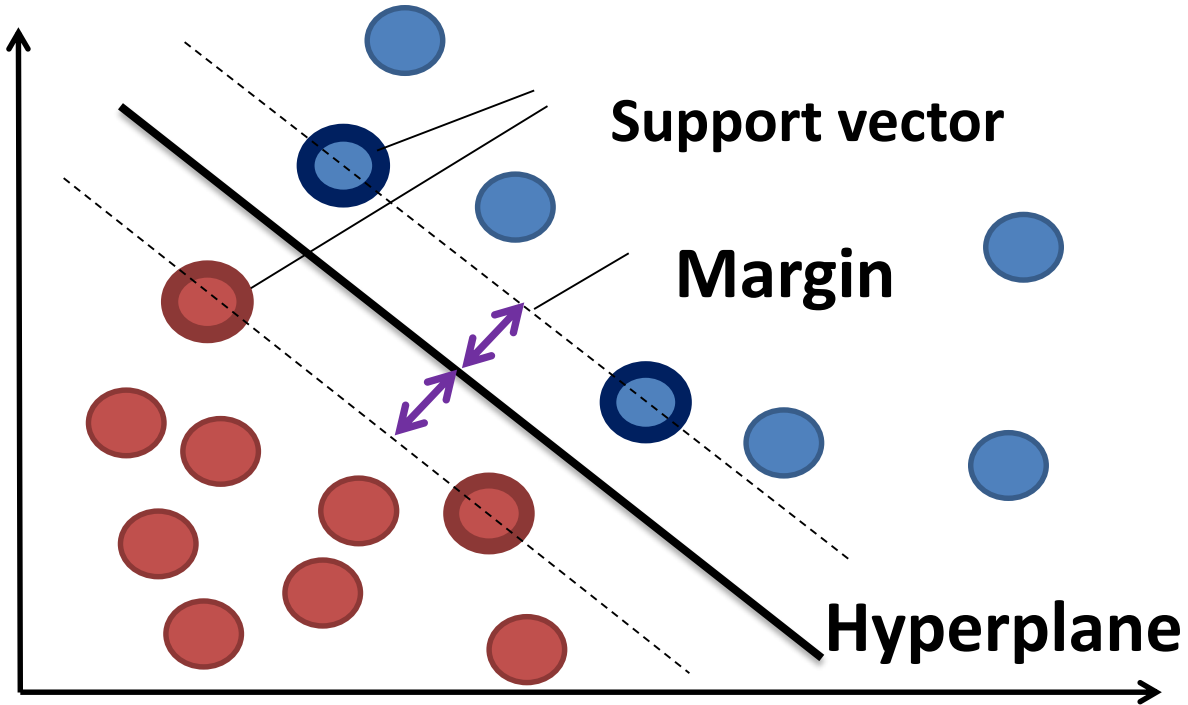


Figure 1. SVM gives a hyperplane which divides two classes.

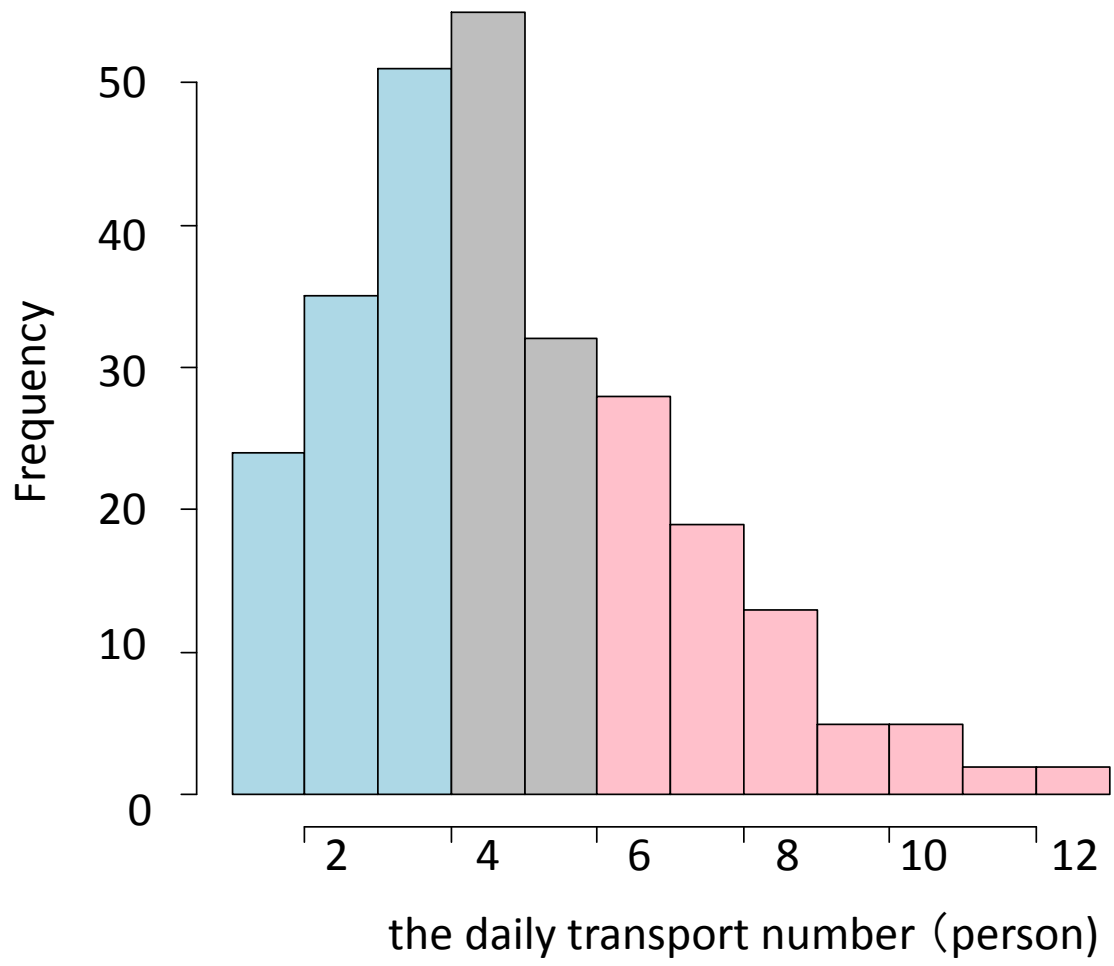


Figure 2. Histogram of daily translation numbers of CI  
Blue means the cases of low risk (the daily transport number of CI is less than 3).  
Red means the cases of high risk ( the daily transport number of CI is more than 6).

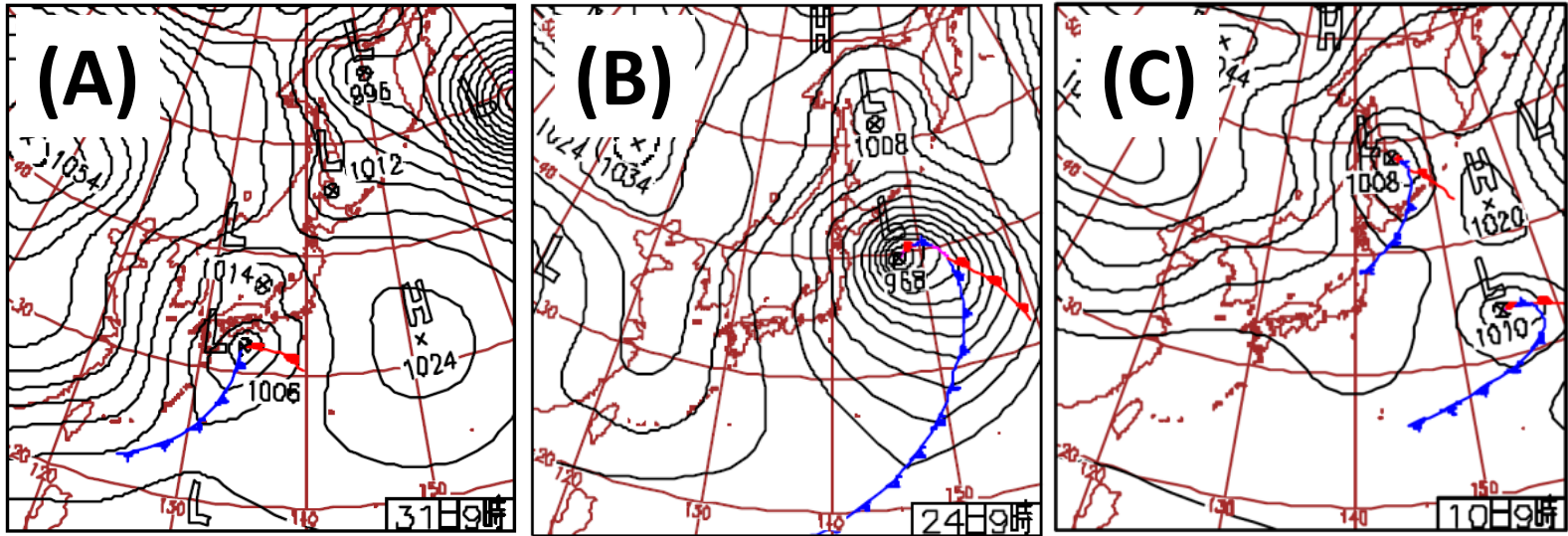


Figure 3. Typical example of three climate patterns are classified

(A): rainy weather

(B): high pressure in the west and low pressure in the east

(C): fine weather

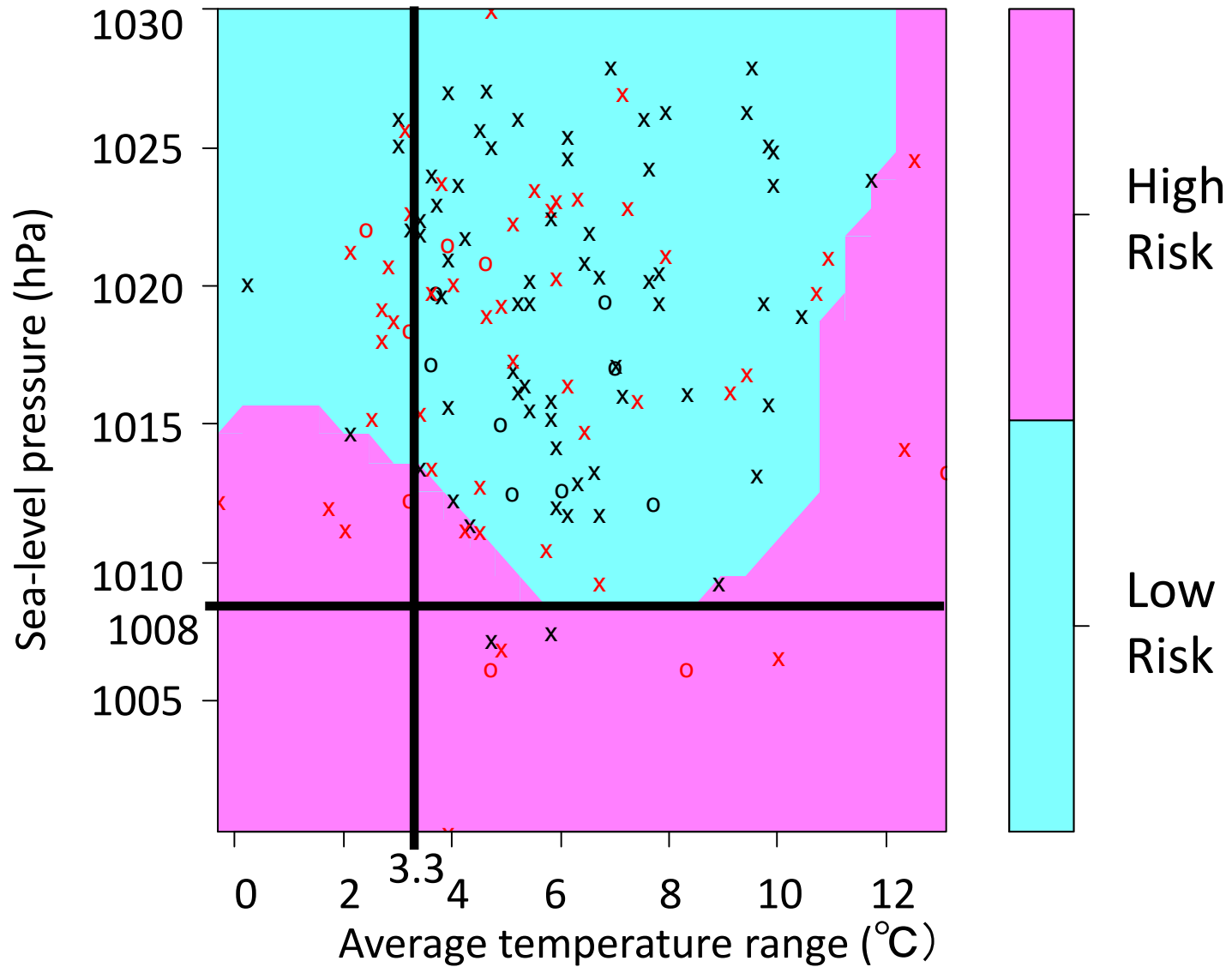
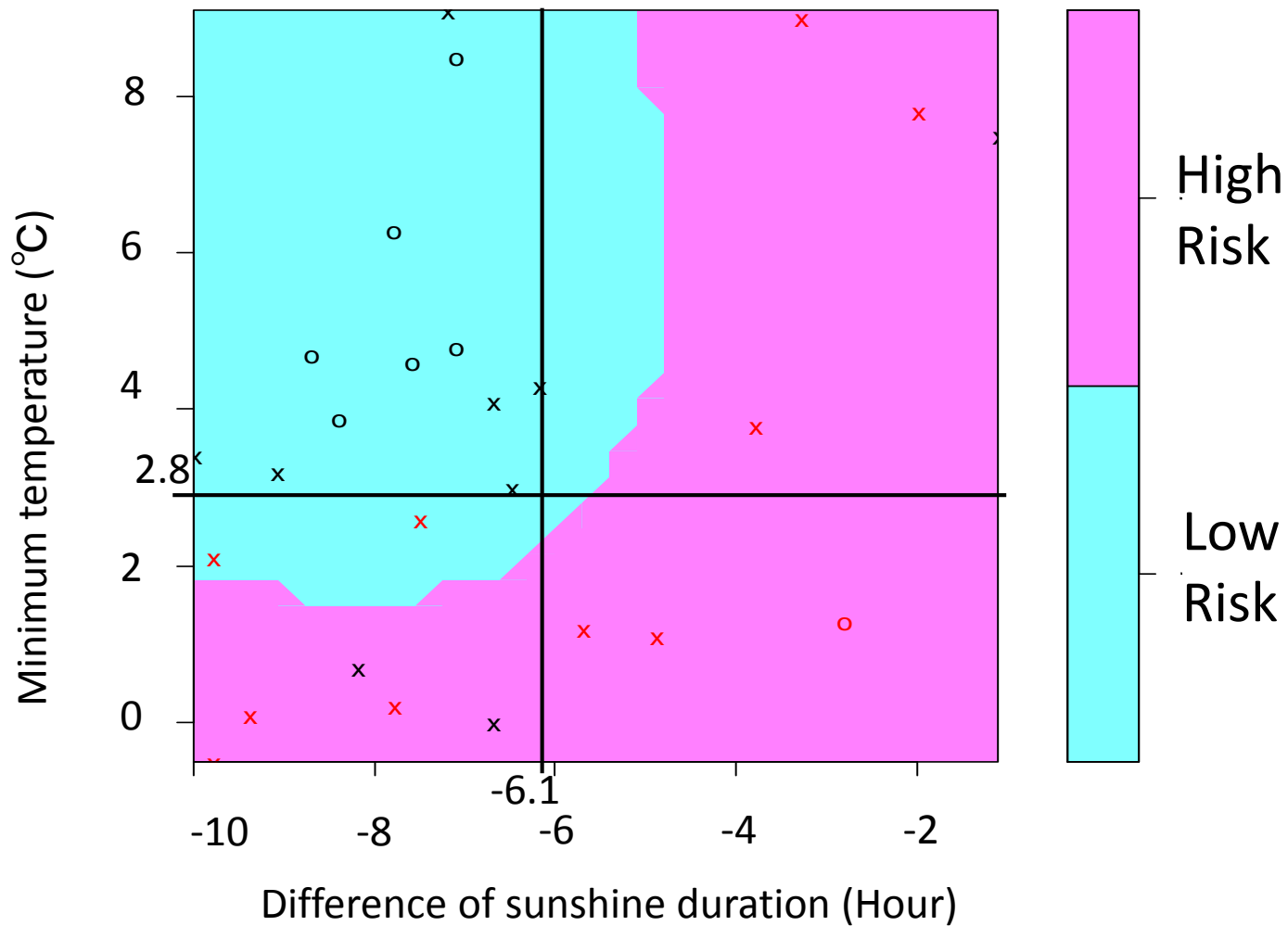


Figure 4. Result of SVM without the classification of climate patterns. The horizontal and vertical straight lines imply the results of decision tree.



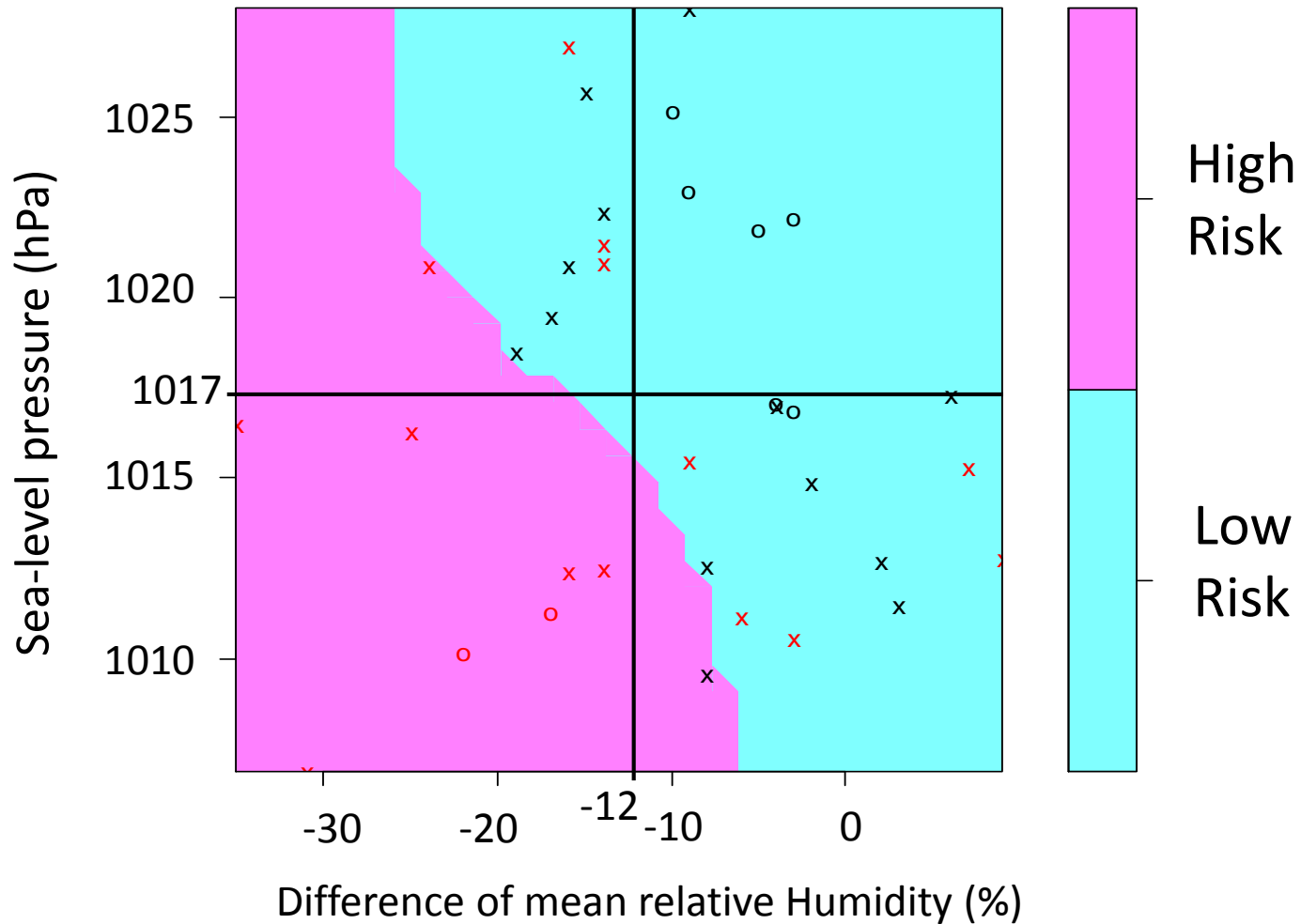


Figure 5-2. The results of classification by SVM for "high pressure in the west and low pressure in the east".

The horizontal and vertical straight lines imply the results of decision tree.

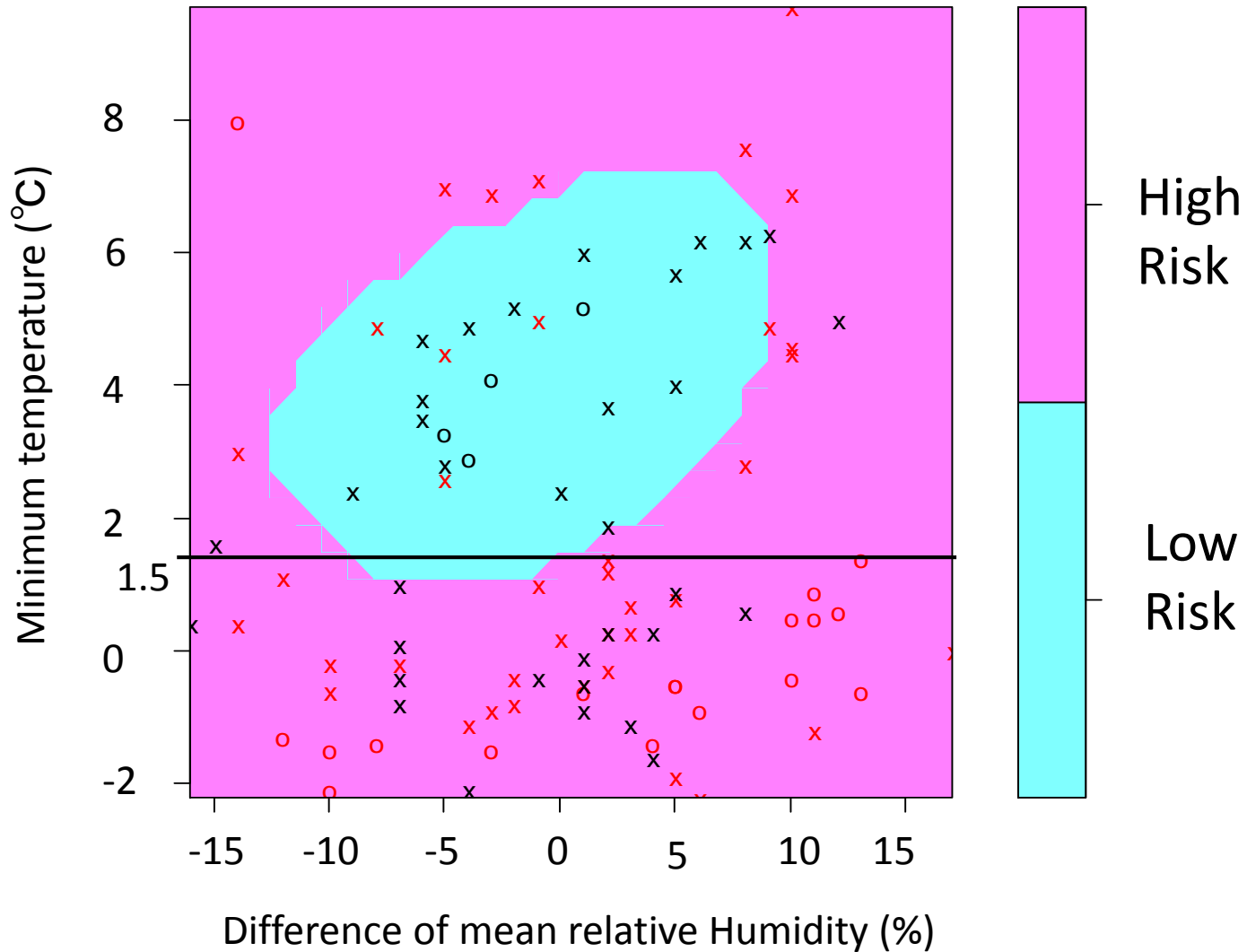


Figure 5-3. The results of classification by SVM for “fair weather”.  
 The horizontal and vertical straight lines imply the results of decision tree.



Table 1. Daily data of 20 climatic elements

---

Sea-level pressure(hPa)	Mean wind speed(m/s)
Mean temperature(°C)	Sunshine duration(hour)
Minimum temperature(°C)	Solar radiation(MJ/m <sup>2</sup> )
Maximum temperature(°C)	Precipitation(mm)
Mean vapor pressure(hPa)	Diurnal temperature range (°C)
Mean relative humidity(%)	(Maximum temperature - Minimum temperature)
Difference of sea-level pressure(hPa)	Difference of mean relative humidity (%)
Difference of mean temperature(°C)	Difference of wind speed(m/s)
Difference of minimum temperature(°C)	Difference of cloud cover (10 degree)
Difference of maximum temperature(°C)	Difference of sunshine duration(hour)
Difference of mean vapor pressure(hPa)	

---