Personalized Spiking Neural Network Models of Clinical and Environmental Factors to Predict Stroke

Document Version
Publisher's PDF, also known as Version of record

Personalized Spiking Neural Network Models of Clinical and Environmental Factors to Predict Stroke

Maryam Doborjeh[1] · Zohreh Doborjeh[2] · Alexander Merkin[3] · Rita Krishnamurthi[3] · Reza Enayatollahi[4] · Valery Feigin[3,5,6] · Nikola Kasabov[1,7]

Abstract

The high incidence of stroke occurrence necessitates the understanding of its causes and possible ways for early prediction and prevention. In this respect, statistical methods offer the "big picture," but they have a weak predictive ability at an individual level. This research proposes a new personalized modeling method based on computational spiking neural networks (SNN) for the identification of causal associations between clinical and environmental time series data that can be used to predict individual stroke events. The method is tested on 804 stroke patients. Given a clinical data set of patients who experienced a stroke in the past and the corresponding environmental time-series data for a selected time-window before the stroke event, the method identifies the clusters of individuals with a high risk for stroke under similar conditions. The methodology involves a pipeline of processes when creating a personalized model for an individual x: (1) selecting a group of individuals Gx with similar personal records to x; (2) training a personalized SNN x model of several days of environmental data related to the Gx group to predict the risk of stroke for x at least one day earlier; (3) model interpretability through 3D visualization; (4) discovery of personalized predictive markers. The results are twofold, first proposing a new computational methodology and second presenting new findings. It is found that certain environmental factors, such as SO_2, PM_{10}, CO, and $PM_{2.5}$, increase the risk of stroke if an individual x belongs to a certain cluster of people, characterized by a combination of family history of stroke and diabetes, overweight, vascular/heart disease, age, and other. For the used population data, the proposed method can predict accurately individual risk of stroke before the day of the stroke. The paper presents a new methodology for personalized machine learning methods to define subgroups of the population with a high risk of stroke and to predict early individual risk of the stroke event. This makes the proposed cognitive computation method useful to reduce morbidity and mortality in society. The method is broadly applicable for predicting individual risk of other diseases and mental health conditions.

Keywords Spiking neural networks · Computational modelling · Personalized modeling · Stroke · Prediction

✉ Maryam Doborjeh
mgholami@aut.ac.nz

1 Knowledge Engineering and Discovery Research Innovation, School of Engineering, Computing and Mathematical Sciences, Auckland University of Technology, Auckland 1010, New Zealand

2 Department of Audiology, School of Population Health, Faculty of Medical and Health Sciences, The University of Auckland, Auckland, New Zealand

3 National Institute for Stroke and Applied Neurosciences, Faculty of Health & Environmental Sciences, Auckland University of Technology, Auckland 0627, New Zealand

4 Department of Engineering Technology, Toi-Ohomai Institute of Technology, Tauranga, New Zealand

5 Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA, USA

6 Research Center of Neurology, Moscow, Russia

7 Intelligent Systems Research Centre, Ulster University, Londonderry, UK

Introduction

Stroke is the second leading cause of death and disability worldwide [1, 2]. Stroke is a neurological condition with a rapid increase of severity of neurological signs within the first minutes and hours after its onset. Early treatment could improve health and well-being outcomes and the success of neurorehabilitation process. Also, stroke is a highly preventable disease, and primary prevention of stroke is the most effective solution to reduce its impact and burden [3]. Thus, stroke risk prediction can contribute both to its prevention and early treatment. There is evidence that theoretically 80 to 90% of stroke can be avoided by modifying various metabolic, lifestyle, and environmental factors, and there are large geographical variations in the population-attributable and lifetime risk of stroke for different risk factors [4, 5].

The high preventability of stroke and population and individual variations in the risk of stroke offers an opportunity for developing systems of stroke occurrence prediction. Numerous studies have been conducted to identify predictors of stroke [2–4]. Such predictors can be a combination of different information sources, including the patient's historical health and medical records, and demographics. Although several investigations have been conducted for the identification of clinical risk factors of stroke, the influences of environmental factors on stroke incidents are not much understood, although these factors may be responsible for up to one-third of stroke burden [4].

Some studies confirmed the relationship between stroke and elevated nitrogen dioxide (NO_2) in Shanghai and Taiwan [6, 7]. Research in China suggested that an enhanced rate of hospital stroke admissions was associated with the effects of different elevated gases including NO_2, sulfur dioxide (SO_2), and O_3. Recent research in the USA reported on the relationships between ischemic stroke risk and particle matter ($PM_{2.5}$) and O_3 exposure, suggesting that a further investigation of pollution and stroke association is essential [8]. Some studies [9–13] explored the effects of stroke risk related to temperature factors and suggested that the rate of stroke occurrence appeared to be higher in colder months during winter-spring. Another study [14] reported that a 2-day environmental temperature measurement period of higher temperatures (the 60 s and 70 s in degrees Fahrenheit) was associated with stroke deaths in selected areas of the USA. Associations of ambient temperature with stroke risk but with a time lag of 3 to 4 days were found in another research [15].

Although several studies focused on the links between single environmental factors and risk of stroke occurrence over the whole studied population [13, 16, 17], modeling of the association between a whole group of different environmental factors and personal health-related features that could contribute to the individualized short-term prediction of stroke is still limited worldwide [18, 19].

The current research proposes a new method to explore how a combination of personal clinical health variables and environmental changes over time can influence the individual risk of stroke from a defined subgroup of the population. For this purpose, we developed a new methodology for personalized predictive modeling using spiking neural networks (SNN), called PSNN. SNN have already been proposed as superior techniques when modeling temporal data, changing over time. SNN represent and learn these changes as sequences of spikes [20]. A class of SNN has been developed to deal with spatio-temporal data [21], such as NeuCube [22, 23] to integrate static and dynamic information [24] and to extract symbolic rules from such data [25, 26]. In this paper, based on available clinical and environmental data, we first define a subgroup of the population at risk, and using this subgroup, we develop a personalized SNN model for each new individual to predict the risk of stroke event before the day of the occurrence. This method supports model interpretability that allows us to recognize which interactions between clinical and environmental risk factors could increase the risk of stroke for an individual or a group of individuals and predict this risk earlier. Compared to the methods proposed in [27] and [28], the current research introduces new methods for personalized modeling of an individual stroke occurrence, as well as identification of combined clinical and environmental risk factors associated with defined clusters of individuals.

Methods

The method introduced here is for the creation of a personalized modeling system to predict individual risk of stroke concerning integrated datasets from clinical data and environmental time series over several days before the stroke. Given a time-window Te of environmental data De and clinical data Dc for patients who experienced a stroke in the past, the method first selects a subgroup of population G for which a personalized SNN model can accurately predict their stroke event at least one day earlier. Then, for every new individual x, (1) a cluster $D_{cg}x$ of individuals from the data set D_{cg} is selected with similar clinical records to the person x; (2) a personalized computational model of SNN x is developed using the environmental data $D_{eg}x$; (3) classifying and predicting the stroke risk for the person after the time-window T_e days; and (4) model interpretability through 3D visualization of the interaction between the changes of the environmental features during the high-risk period for this person.
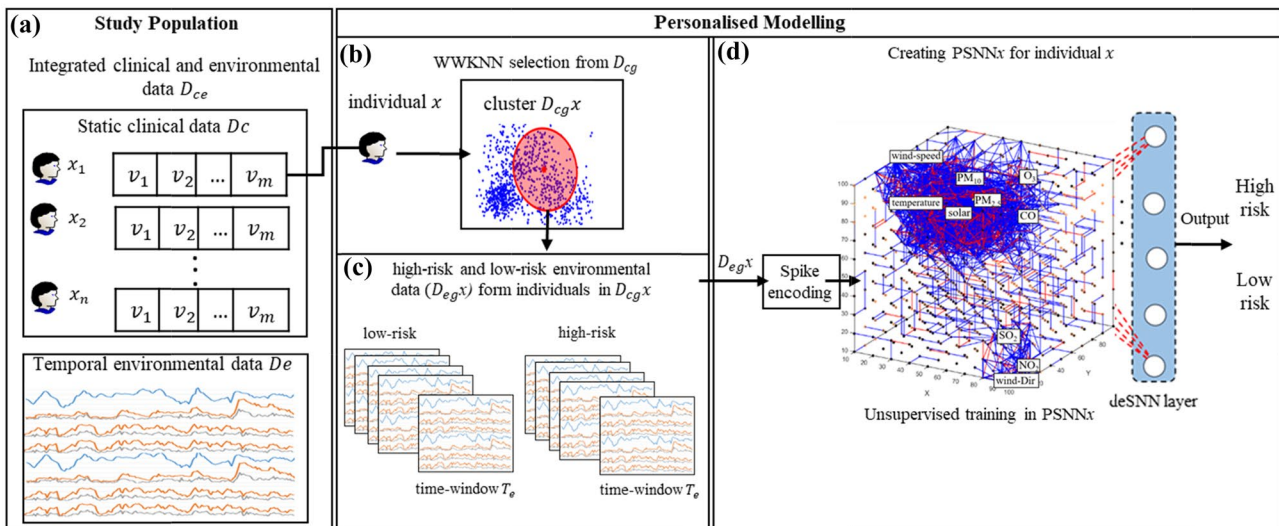
Fig. 1 Schema of the personalized modeling system for integrated clinical data and dynamic environmental data D_{ce} (shown in **a**) for individual stroke prediction. (**b**) For a new individual x, a cluster $D_{cg}x$ of individuals is selected from a data set D_c of patients with stoke in the past in respect to similarity in their clinical data. (**c**) A time-window T_e of several days of high-risk and low-risk environ-mental data changes prior to stroke event of each patient from the cluster $D_{cg}x$ are extracted, called $D_{eg}x$. (**d**) Selected time series $D_{eg}x$ are used to train a PSNN x model. The model is then tested using the high-risk and low-risk environmental periods from individual x to detect if the person x is in a high- or low-risk period for a stroke occurrence

Method and System for Personalized Predictive Modeling on Integrated Personal Clinical Data and Dynamic Data of Environmental Changes

The architecture of the proposed methodology is illustrated in Fig. 1, which represents the computational steps of building a personalized predictive model for an individual.

Figure 1b shows that for a new individual x, the k nearest neighboring samples is found by computing a pairwise normalized Euclidean distance between the clinical health information (one static vector) of individual x and the other individuals' clinical records. We also included the importance of the data features when computing the distance. This was measured by signal-to-noise ratio (SNR) [29] that is a statistical measurement to rank the variables with respect to their power in differentiating the samples to different classes (health conditions). This method of selecting the nearest samples to the individual x is called weighted–weighted distance k-nearest neighbors (WWKNN) method [28], where the first W is the SNR rank of the variables and the second W is the Euclidean distance. Figure 2a illustrates the distance between clinical records of one randomly selected individual x (id-1 among 804 patients) and the other 803 individuals. The green bars are those individuals with high similarity to individual x when an adaptive radius threshold r is applied (formed cluster $D_{cg}x$) to define the neighborhood radius. We assigned three different values to the threshold r which are μ or μ + σ or μ + σ to optimal the value of k, where μ is the mean value and σ is the standard deviation computed in the Euclidean distances of all individuals' data vectors to individual x vector.

For each of the k selected individuals in $D_{cg}x$, the time in which an individual had a stroke is indexed in the environmental data. When moving backwards from the index time, the closer an individual is to the onset of stroke occurrence, the greater interaction of risk factors is likely to be observed. Therefore, a time-window (in our experiment here, the time-window T_e has a length of 7 days = 168 h) positioned before the stroke onset can be considered as a "high-risk" interval. Another 7-day time-window positioned at 2 months before the stroke can be considered as a "low-risk" interval. Figure 1c shows that for every individual from $D_{cg}x$, two environmental intervals are extracted as two temporal samples, one belongs to the class "high-risk" environment and the other one belongs to the class "low risk" environment. Figure 2b shows an example of three environmental variables changing over a time-window of 168 h from two classes: high-risk and low-risk environmental data. The method allows to explore different lengths of the time-window T_e, and for each time-window, different subgroups of individuals can be selected for which the environmental factors in this window in combination with their clinical factors can cause a high risk for stroke after the selected number of days.

Figure 1d shows that the selected environmental data samples $D_{eg}x$ are used to build a PSNN x, model for individual x for mapping, learning, visualizing, and classification of "high-risk" and "low-risk" environmental data periods. The proposed PSNN x model is a reservoir computing system
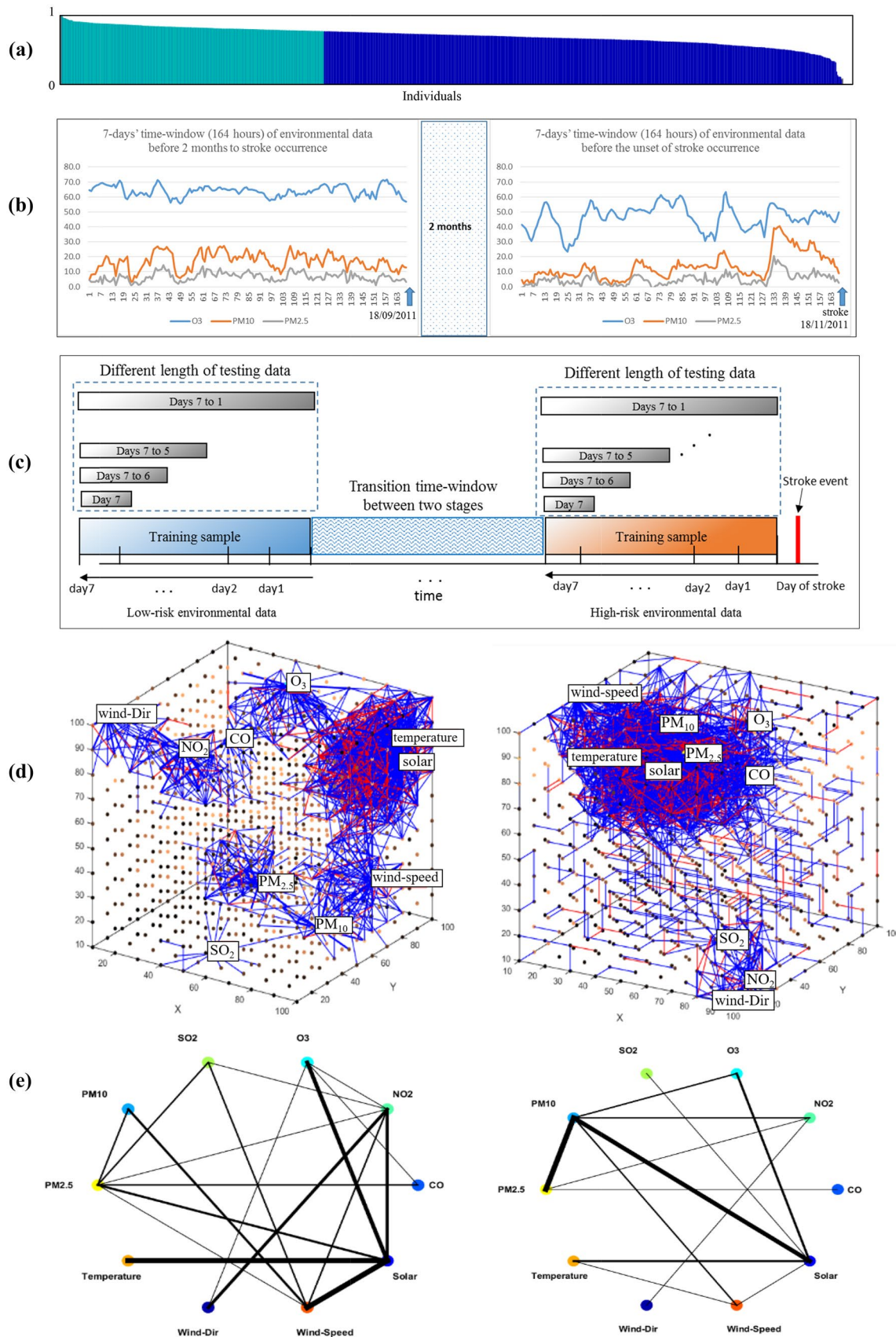
◀**Fig. 2** (**a**) A cluster of similar individuals to individual x in terms of clinical data are highlighted in green, referring to those samples who are withing a neighborhood radius threshold r around individual x, where r is an adaptive threshold for every personalized model ($r = \mu + \sigma$) which resulted the best accuracy of classification between high risk and low risk. (**b**) An example of the environmental samples related to an individual (from the green bars) with 3 features (O_3, PM_{10}, and $PM_{2.5}$), where left is a 7-day data (164 h) from "low-risk" and right is from "high-risk." (**c**) The design of the training and testing datasets for creating PSNN models. The training samples have a fixed length (7 days), while the length of the testing samples is changing from a 7-day period to 1-day period (prior to stroke) to identify the best early prediction timepoint for this individual possible stroke occurrence. (**d**) The trained PSNN models with the low-risk environmental period (left) and high-risk environmental period (right). (**e**) The feature interaction networks in the two PSNN models for low-risk and high-risk environmental periods

that consists of artificial spiking neurons as processing elements, spatio-temporal connections between the neurons, and biologically plausible algorithms for learning from data [23, 30–32]. Here, the designed PSNN x model is a recurrent network which is transpired as a promising architecture to learn spatio-temporal patterns from spatio-temporal data [23]. Modeling of environmental samples using PSNN comprised the following phases:

- Encoding of environmental samples to spikes.
- Spatial mapping of the environmental features into a 3-dimensional PSNN model.
- Unsupervised learning in the PSNN model.
- Supervised learning to detect the association between the training samples and their class labels (high-risk and low-risk environments). Then, the environmental samples of individual x (which were excluded from the learning phase) were used to cross-validate the model.
- Optimization process.

The aforesaid methodological phases are explained as follows:

Encoding of Environmental Time-Series Data To transfer the temporal samples into an SNN model, they need to be first encoded into sequences of binary events, called spikes which represent significant changes in time. For this, a threshold-based representation method (TBR) method (examples shown in [33–44]) is used to encode the environmental data changes to spikes (encoded to 1 if an upward change exceeds a pre-defined encoding threshold, or to −1 for a downward change).

Environmental Data Mapping into a Personalized SNN Model In this dataset, the environmental data samples are defined using 10 environmental time series variables. To spatially map these variables, we first created a

3-dimensional PSNN model which contains 1000 artificial spiking neurons as computational units. The temporal variables are mapped to the PSNN model, so that the closer the variables are mapped together, the higher the correlations between their encoded spike sequences [45, 46]. When the spatial information of the samples is mapped, the PSNN connectivity is initialised using the small-world-connectivity rule (SW) [23].

Unsupervised Learning in the PSNN Model To learn the "deep in time" spatio-temporal relationships between the temporal environmental variables, we used an extension of Hebbian learning rule, called spike-timing dependent plasticity (STDP) [20]. The STDP rule is a neuroscientific concept that represented an increase in synaptic efficiency which is driven by a presynaptic neuron that repeated stimulation of a postsynaptic neuron. The STDP learning modifies the PSNN connectivity according to the relative timing of the pre- to post-synaptic spikes. If two neurons i and j are connected, *wij* increases if neuron i fires first and then neuron j within a defined time interval. On the other hand, *wij* decreases if neuron j fires first and then neuron i. It means that *wij* describes the temporal relationship between neuron i and j with respect to the time of spiking. In this case, whole spatio-temporal associations and patterns across the environmental variables, rather than single variable, are learned as triggering factors for a stroke event.

Supervised Learning, Classification, and Prediction When the unsupervised learning process with the training samples is completed, the training samples are used again for supervised learning in an output dynamic evolving SNN (deSNN) classifier [21]. This procedure learns the association between the trained patterns in the PSNN model and output class label information (e.g., high risk vs low risk). Figure 2c shows the length of the temporal environmental samples for training and testing phases. A time-window of 7-day (168 h) length (can be adjusted by end-users) before the stroke is defined to form the training dataset which contains several individuals' samples. Then, the 10 environmental features are mapped into a 3D PSNN model and an unsupervised learning algorithm [20] is used to capture the spatio-temporal relationships between the features over 7 days in both low-risk and high environmental periods (Fig. 2d-left and 2d-right). The causal temporal interactions between the 10 environmental variables over the selected T_e periods of 7 days are shown in Fig. 2e which demonstrate how the changes in one feature influenced the other features on the following day. The trained PSNN models are later tested with a smaller length of the testing samples (not used for training) to validate the ability of the system for early prediction of stroke occurrence.
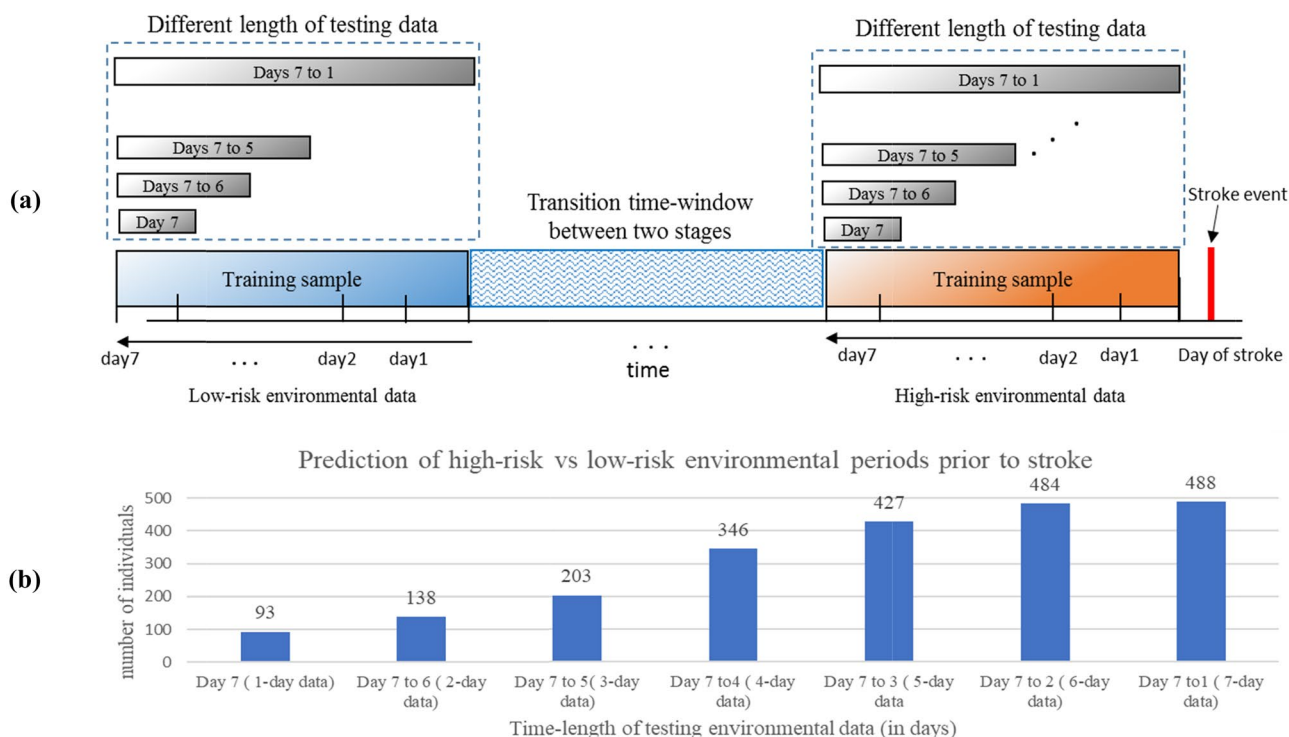
Fig. 3 (**a**) The design of the testing data (environmental time-series in our case from 7 days to 1 day of data). (**b**) The PSNN models differentiated the "high-risk environment" vs "low-risk environment" for 488 individuals when tested with 7 days of environmental data prior to stroke occurrence. This indicates that there is an association between the 7-day environmental changes and the risk of stroke occurrence for a subgroup of 488 individuals in the whole population. The number of individuals with the correct prediction of low-risk environmental period (risk of stroke) was reduced when the length of the testing environmental time-series was shortened from 7 days to 1 day

Study Population and Datasets

Data involved clinical health records from patients ($N = 804$) who had stroke occurrences between 1st March 2011 and 1st March 2012. There were 382 (47.5%) females with the mean age of 71.11 and 422 (52.4%) males with the mean age of male $= 69.75$. Each patient's data includes 37 static features such as age, gender, ethnicity, blood information (cholesterol, pressure), stroke history, disease history (diabetes, migraine, epilepsy/seizures, etc.), heart disease (heart attack, irregular pulse, and failure).

Environmental data were recorded over the same period (1st March 2011 to 1st March 2012) by 10 meteorological monitors positioned in Auckland city, New Zealand. The measures included the following: carbon monoxide (CO), nitrogen dioxide (NO_2), ozone gas (O_3), sulfur dioxide (SO_2), and particulate matters (PM_{10} refers to an aerodynamic diameter smaller than 10 μm and $PM_{2.5}$ refers to particles with an aerodynamic diameter smaller than 2.5 μm), temperature (°C), wind-direction average (°),[1] wind-speed

(m/s),[2] and solar radiation (W/m²).[3] The data were recorded on an hourly basis; therefore, 8784-time points were measured over the 1 year.

Results

To model the differences between the patterns of low and high risk of environmental data for each person, personalized models were created separately for 804 individuals from the data set. Each PSNN x model of a person x was trained in our experiment with a time-window Te of 7-day environmental data of a group of k nearest neighboring individuals to this person (selected using WWKNN method) and then was tested 7 times using different lengths of the environmental samples from i (testing data length varied from 7-day period to 1-day period, prior to stroke occurrence). Figure 3

---

[1] Wind direction is measured in degrees clockwise from due north (measured in units from 0° to 360°). Consequently, a wind blowing from the north has a wind direction of 0° (360°); a wind blowing from the east has a wind direction of 90°; a wind blowing from the

Footnote 1 (continued)

south has a wind direction of 180°, and a wind blowing from the west has a wind direction of 270°.

[2] Meters per second.

[3] Watts per square meter.

depicts that when PSNN models were tested with 7-day environmental samples prior to the stroke, the high-risk and low-risk samples were correctly classified for 488 individuals. However, the number of individuals reduced when the PSNN models were tested using a smaller time-length (a 6-day to 1-day period) for prediction of stroke occurrence on the 7th day. The findings in Fig. 3 suggest that this subset of 488 individuals' models showed associations between 7-day environmental data changes and their risk of stroke, forming a subgroup of individuals G. Our hypothesis is that every new individual who has similar clinical variables to the population G of individuals can benefit from a PSNN to predict their stroke risk using 7 days of environmental data. For the rest of 804–488=316 individuals, other suitable PSNN models should be explored, using a larger window Te of environmental data (e.g., 8, 9, 10, …,20 days as suggested in [47]). Here, for each time-window, a separate subgroup of individuals can be identified that associates their clinical variables with the environmental variables during this time-window. We have studied what clinical variables define the subgroup G of 488 individuals for which 7 days of environmental variables can be used to predict their risk, in contrast to the rest 316 individuals. This study is important for the future applicability of the proposed method in clinical practice.

As stated earlier, every PSNN model was tested 7 times using different lengths of the environmental period prior to the stroke; hence, among these 488 individuals, a subset of individuals whose high-risk environmental periods were detected correctly in at least 4 rounds out of these 7 testing rounds (e.g., 1,2,3 and 4 days before the stroke) was selected as a group of strongly affected patients by current environmental changes. This subset represents those individuals who experienced the effect of causal interactions in longitudinal environmental time-series with their personal, clinical data that contributed strongly to increasing their risk of stroke. As a result, 169 individuals were selected for further quantitative analysis of their PSNN models. Therefore, the whole 804 individuals were categorized into two groups: (1) the affected group (AG) of 169 patients (accurate prediction of at least 1, 2, 3, and 4 days before the stroke) and (2) the non-affected group (NAG) of 635 patients.

To identify the between-group differences, we analyze the distribution of the patients (in percentage) in the affected and non-affected groups with respect to their family health history (Fig. 4a) and their personal health history (Fig. 4b). Figure 4c represents the differences in the mean value of some clinical health features in the AG vs NAG.

Our findings suggest that the risk of stroke in the studied population was associated with certain environmental changes when the individuals belonged to a defined cluster of the following clinical risk factors: a family health history factors (stroke in family, diabetes in the family; depicted in Fig. 4a); personal health history, high cholesterol, vascular/heart disease (depicted in Fig. 4b); and greater values in age, weight, and blood pressure (depicted in Fig. 4c).

To investigate how the interactions between environmental variables during the chosen time-window of 7 days before stroke affected an individual risk of stroke, we built personalized models for each of these 169 patients to capture the within-group differences of high-risk vs low-risk environmental periods. Here, for every individual $x = \{1, \dots, 169\}$, we selected a cluster of patients using the WWKNN method concerning their clinical data similarity. The size of the selected cluster is different for each of these 169 individuals, depending on the density of the similar individuals in the neighborhood radius. Figure 5 plots the number of k similar samples to each of these 169 individuals, selected for building 169 PSNN models. Each created PSNN model was trained with two sets of environmental time-series (from high-risk and low-risk classes) that belong to the k nearest individuals to an individual x. These environmental time-series were encoded into spikes to demonstrate certain upward and downward changes in the values of environmental features over 7-day periods in both high and low-risk intervals.

Figure 6a depicts the average of positive and negative spikes derived from the 7-day environmental data in high-risk samples. This represents that in the high-risk environment, the values of CO, NO_2, O_3, SO_2, PM_{10}, and $PM_{2.5}$ have been increasing more than decreasing, therefore, generating more positive spikes than negative. On the other hand, the values of temperature, wind-speed, wind-direction, and solar radiation, which are inter-related climatic conditions, have been decreasing more than increasing. These patterns demonstrate the associated environmental changes over 7 days before stroke occurrence that influenced the risk of stroke for these 169 affected patients in Auckland in 2011–2012. Except for O_3, the mentioned pollutants are mainly generated because of burning fossil fuels. The presence of NO_2 and SO_2 together with water and oxygen will result in the production of nitric, nitrous, and sulfuric acids. Particulate matters (PM), especially $PM_{2.5}$, due to their small size can penetrate the lungs, which triggers respiratory diseases [48]. These particles can also enter the blood circulation system that may lead to chronic diseases and cause vascular inflammation and hardening of arteries that may result in ischemic stroke or heart attack [49–51]. Our findings in Fig. 6a are in alignment with the literature that suggested $PM_{2.5}$ as a risk factor of stroke occurrence [49, 52]. Figure 6a also reported an association between the ozone (O_3) increase and the high-risk period of stroke occurrence. Ozone sis an allotrope of oxygen that can be generated by short wavelengths of the ultraviolet spectrum, particularly UV-C (200–280 nm) and vacuum UV (100–200 nm) [53]. Ozone was seen to alter blood
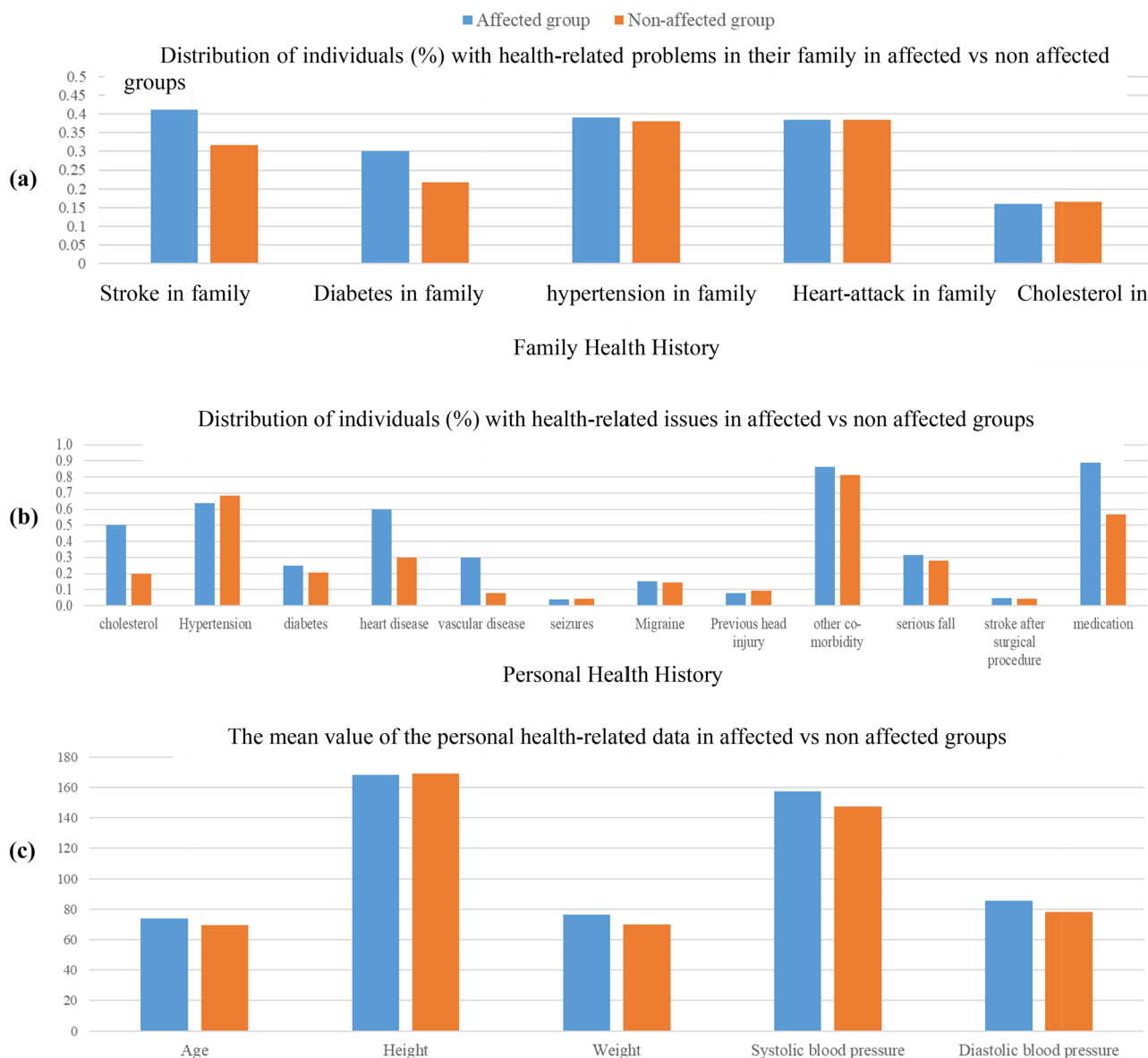
Fig. 4 Clinical records of patients in two groups: affected vs non-affected groups by environmental changes. (**a**) The number of individuals with a history of health issues in their family records shows that the most of them had family members who had a stroke in the past; (**b**) the number of individuals with history of a health issues in their personal, clinical records shows a higher level of cholesterol, diabetes, vascular/heart disease, comorbidity, serous full, and medication for the affected group; (**c**) the mean value of the last measured personal, clinical health variables show a greater values in features age (over 65), weight, systolic blood pressure (over 155 mm of mercury—mmHg), and diastolic blood pressure (over 80 mmHg) for the affected group

coagulation mechanism and cause irregular heart rate and systemic inflammatory responses [54, 55] and hence was reported in the literature to be in association with stroke occurrences [56, 57].

The encoded spikes from 7-day environmental data were used as input data for training PSNN models. The environmental features were mapped into a 3D PSNN model that topologically preserves the temporal differences of the data features. This is performed by computing the correlation between the spike trains of all the 10 environmental features.

The most correlated features are mapped to closer input neurons inside the PSNN.

For each of the 169 individuals in the affected group, we developed two separate PSNN models to map and model the temporal environmental changes of the high- and low-risk periods and study the differences. The PSNN models were spatially mapped into the 3D space of spiking neurons and trained environmental time-series. The mapped PSNN models learned the temporal associations "hidden" between the environmental features during the unsupervised STDP
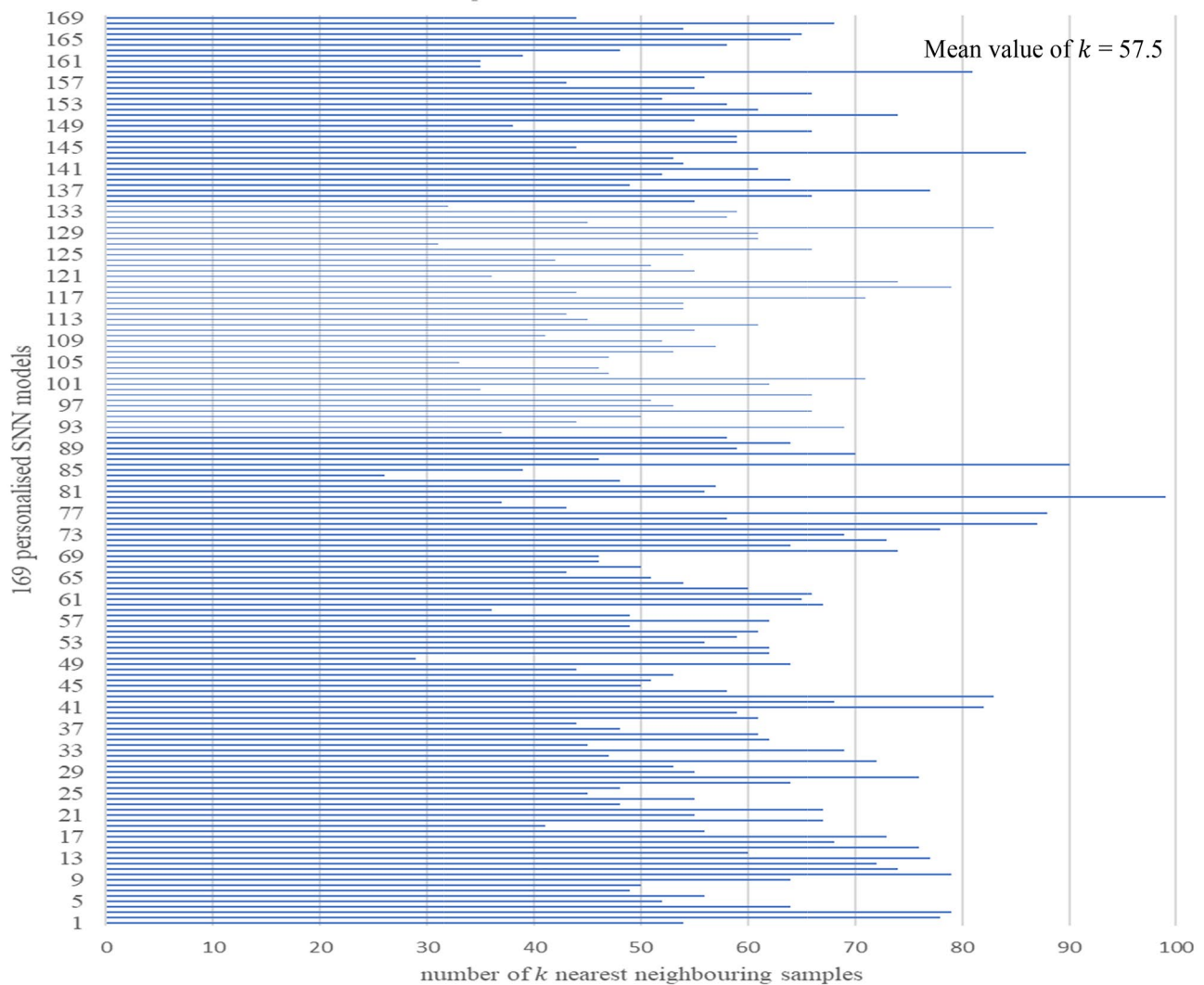
Fig. 5 For each of the personalized models of individuals $x = \{1, .., 169\}$ (shown on the y-axis), k neighboring samples selected with respect to a neighborhood radius r which is an adaptive threshold ($r = \mu + \sigma$) and is a different value for each personalized model. This led to select an optimal value for k in each personalized model (k is shown on the x-axis) and on average, $k = 57.5$

learning algorithm [20] while learning from 7-day data. Figure 6b shows the level of causal interactions that each environmental feature has with other features during the 7 days, averaged across all the 169 PSNN models in high risk (red) vs low risk (blue). This shows a greater causal interaction in high-risk than the low-risk period reflecting the associated environmental risk factors.

When the PSNN models are learning from environmental data using the unsupervised STDP learning algorithm [20], the spatio-temporal relationships between the features are formed as weighted connections.

Figure 7 illustrates the absolute value of positive and negative connection weights in the PSNN models of 169 individuals, trained by high-risk (in a) and low-risk (in b) environmental data. By comparing Fig. 7a and b, the absolute

value of connections is higher in the high-risk period than in the low-risk period. It may suggest that frequent fluctuations in environmental features might be considered as external risk factors to increase the risk of stroke occurrence. For statistical analysis, we extracted the quantitative information of the connection weights from 169 patients' PSNN models of high-risk and low-risk environments and used ANOVA to measure the t-test p-values as reported in Table 1.

Personalized Profiling of Individual Risk of Stroke Using Environmental Data

The study of interactions among environmental variables over time, related to personal data before stroke occurrence, is a challenging task as several variables can influence the
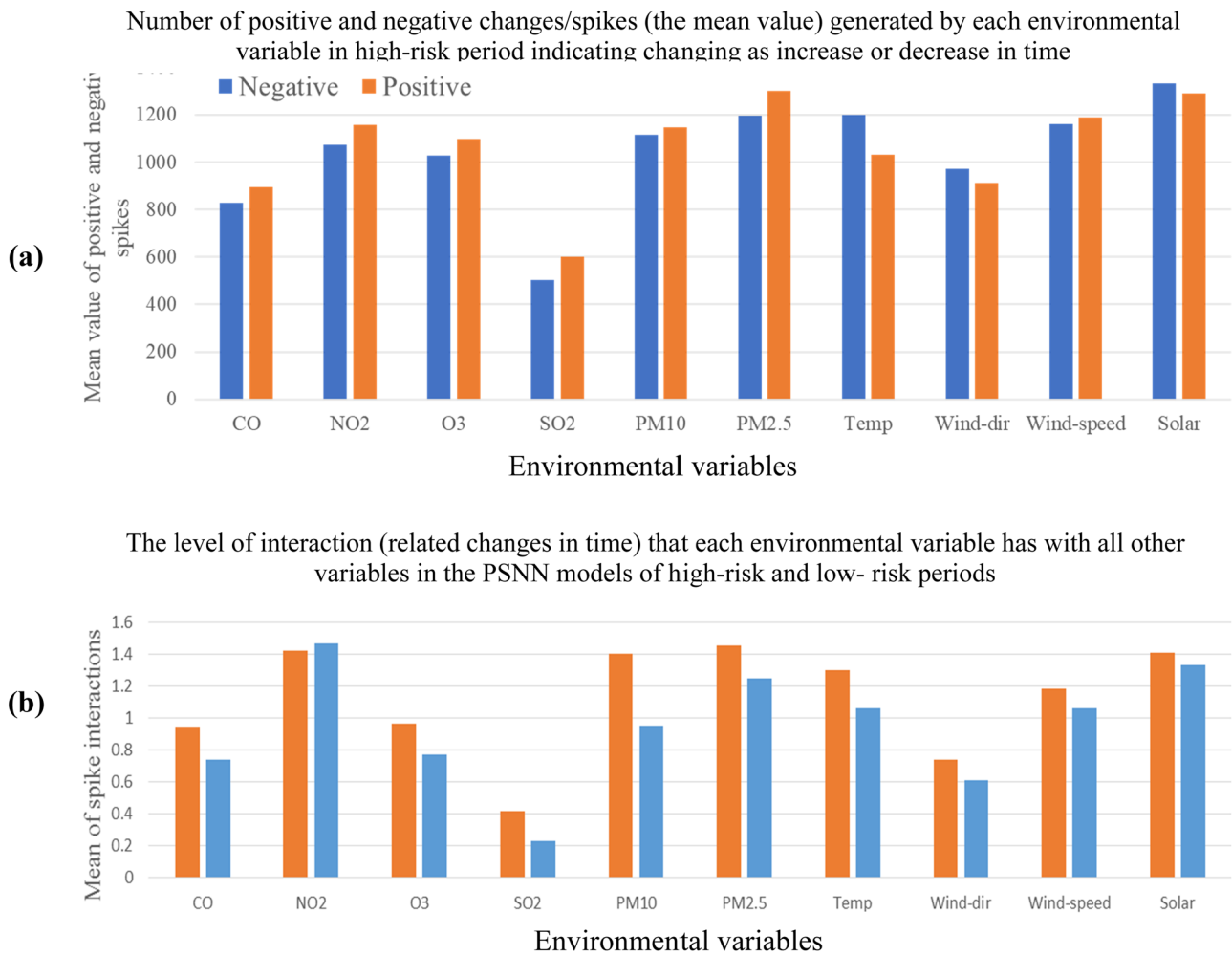
Number of positive and negative changes/spikes (the mean value) generated by each environmental variable in high-risk period indicating changing as increase or decrease in time

(a)

The level of interaction (related changes in time) that each environmental variable has with all other variables in the PSNN models of high-risk and low- risk periods

(b)

Fig. 6 (**a**) The number of positive and negative spikes (mean values) related to the increases and decreases in environmental time-series for the high-risk period, averaged across all the 169 individuals. (**b**) The level of influence (causal relationship) that one variable has on the others over 7 days of high-risk (in orange color) and low-risk (in blue color)

other ones, either directly or indirectly. Here, the proposed personalized modeling method and system offered a capable and explicable profile of an individual to explain the relationships between environmental variables that potentially increased an individual's risk of stroke for a person or a group of persons. Using the proposed PSNN method and system, we can create a personalized profile for each person that results in an improved understanding of personal factors

The sum absolute value of positive and negative connection weights
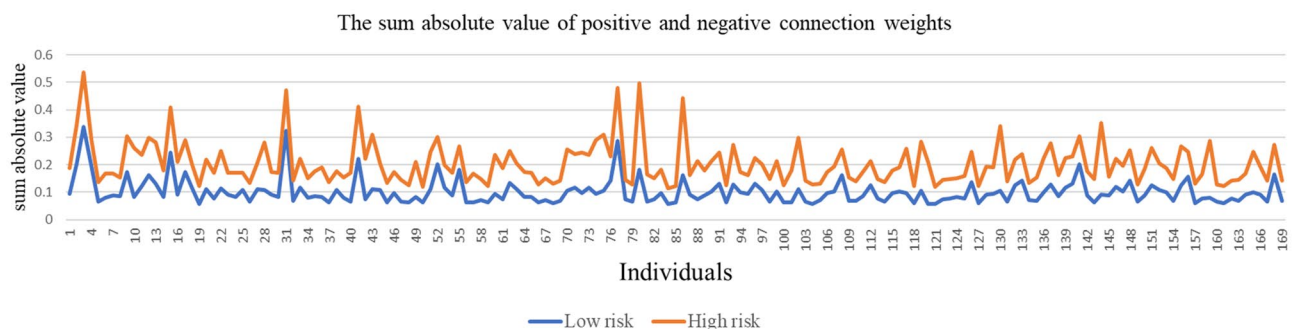
Fig. 7 The sum absolute value of positive and negative connection weights in each of the trained PSNN models (for 169 patients) in high risk (red) vs low risk (blue)

Table 1 A *t*-test *p*-value demonstrates the significant difference between the level of interactions for each environmental variable across 169 patients' models in high risk vs low risk. Variables SO_2 and PM_{10} have shown the lowest *p*-values followed by CO and $PM_{2.5}$ variables, representing the most important variables for discriminating the two groups

Variables	CO	NO2	O3	SO2	PM10	PM2.5	Temperature	Wind-Dir	Wind-Speed	Solar
t-test p value	0.0002	0.5	0.001	3.10E-18	2.20E-10	0.009	0.001	0.0017	0.07	0.3

that increased the risk of stroke. Figure 8a represents the PSNN models (trained by high-risk and low-risk environmental time-series) of a 21-year-old (female) patient who had a stroke on 18 Nov 2011 in Auckland, NZ. The PSNN models demonstrated that the spatio-temporal relationships between the environmental variables are different in high-risk vs low-risk environments for this patient with the following conditions: epilepsy, head injury, migraine, and family history of heart attack, hypertension, and diabetes.

The amount of spatio-temporal interactions between these environmental variables (shown in Fig. 8a) is measured by a feature interaction network (FIN) graph, illustrated in Fig. 8b. For this patient, the FIN graph of high risk represents large interactions between variables NO_2, wind-direction, and $PM_{2.5}$; variables PM_{10} and $PM_{2.5}$; and variables O_3, solar, SO_2, and temperature which explain how the changes in some features influenced the changes in other features over 7 days before the stroke. On the other hand, different level of interaction was measured in the low-risk environmental period for this patient. These findings are personalized and can be different for another patient, suggesting that the proposed PSNN modeling is a promising approach of capturing individual characteristics that can potentially lead to customization of healthcare, decision-making, treatments, and practices as the models are being tailored to individual information.

Figure 8c shows that the data from high-risk and low-risk environmental periods demonstrated different activated areas (shown in %) around each environmental feature in the PSNN models. A larger activated area around an environmental feature refers to stronger influential changes in the value of this feature during the 7 days of high-risk (Fig. 8c-left) and low-risk (Fig. 8c-right) environments. This refers to important environmental markers in increasing the risk of stroke occurrence for an individual.

Figure 9 presents the personalized profiles of another two randomly selected patients from two clusters of subjects with the following information: age > 70, a family history of stroke, high cholesterol, diabetes, vascular/heart disease. These patients had a stroke on 21 Apr 2011 and 30 Jan 2012 respectively in Auckland, NZ. The models were separately trained with 7-day data of high-risk environmental periods related to KNN individuals to these patients. The right-side graphs show the temporal/causal interactions between the environmental features as important measurements for the identification of environmental changes that influenced the risk of stroke.

Figure 9a demonstrates great interactions between PM_{10} and $PM_{2.5}$ and NO_2, also, between the temperature, solar, and wind-speed during the 7 days in the high-risk period. Figure 9b illustrates great interactions between PM_{10} and $PM_{2.5}$, also, between the temperature, solar, and O_3 during the 7 days in the high-risk period.

Discussion

The findings, obtained with the use of the prosed personalized modeling methodology, suggest an association between the occurrence of stroke and changes of environmental factors over 7-day period prior to the stroke event in a group of individuals with particular characteristics, the so-called an affected group (AG) for this time-window period. These individuals have the following demographic and clinical risk factors: a family history of stroke diabetes and hypertension (depicted in Fig. 4a); a personal history of a high level of cholesterol, diabetes, obtained with the proposed vascular/heart disease, serious fall (depicted in Fig. 4b); older age (over 65); and overweight and obesity (depicted in Fig. 4c). The difference in distribution by gender suggests the effects of environmental changes were 10% more noticeable on males than females. Participants in the AG were older; however, females and males in the AG were of similar ages. For an individual in the AG with the aforementioned factors, the risk of stroke was increased by certain patterns of 7-day environmental changes (prior to stroke onset) that includes increment in CO, NO_2, O_3, SO_2, PM_{10}, and $PM_{2.5}$, and decrement in wind-speed, temperature, and solar. Our findings in Fig. 6 imply greater interactions between the environmental features in a high-risk period (the 7 days before the stroke occurrence) than a low-risk period (the 7-day period positioned at 2 months prior to the stroke event). This indicates that there were causal relationships between changes in the values of environmental features during the 7-day period that increased the risk of stroke.

Hitherto, numerous studies have been undertaken to explore clinical risk factors of stroke [4, 58, 59]. However, little research has been conducted to analyze the effects of environmental factors on stroke occurrence [13]. Some studies to date discovered associations between some seasonal environmental patterns and stroke incidences [9–13]. For instance, the rate of stroke occurrence appeared to be diverse as a function of environmental temperature [14, 15]. Some studies in China revealed the associations between stroke

High-risk   Low-risk

(a)

(b)
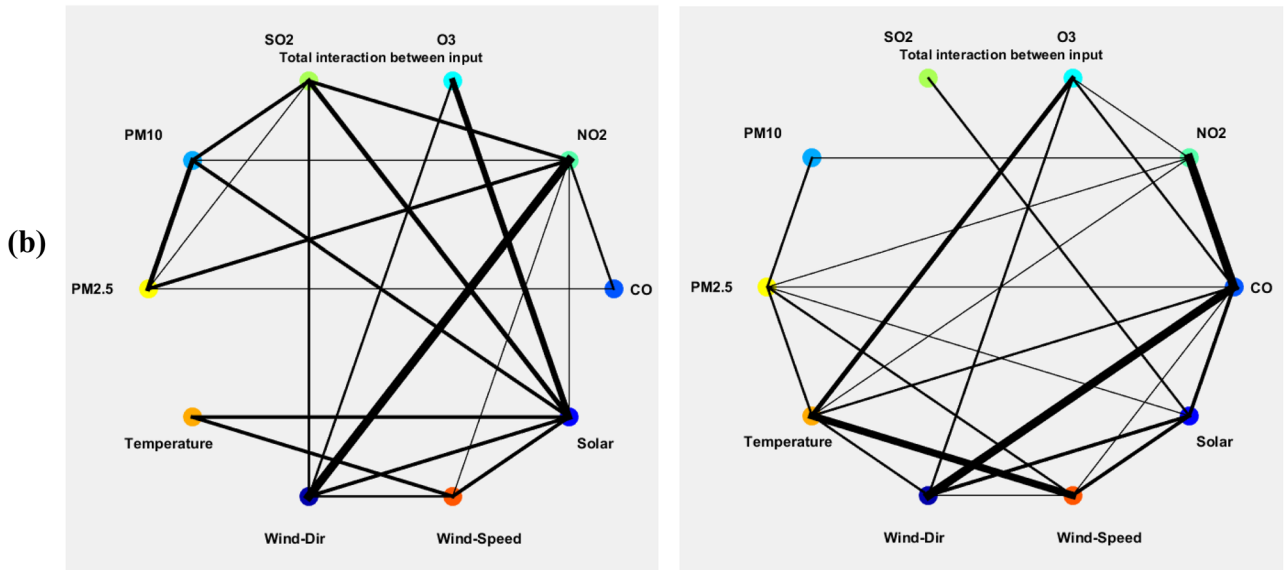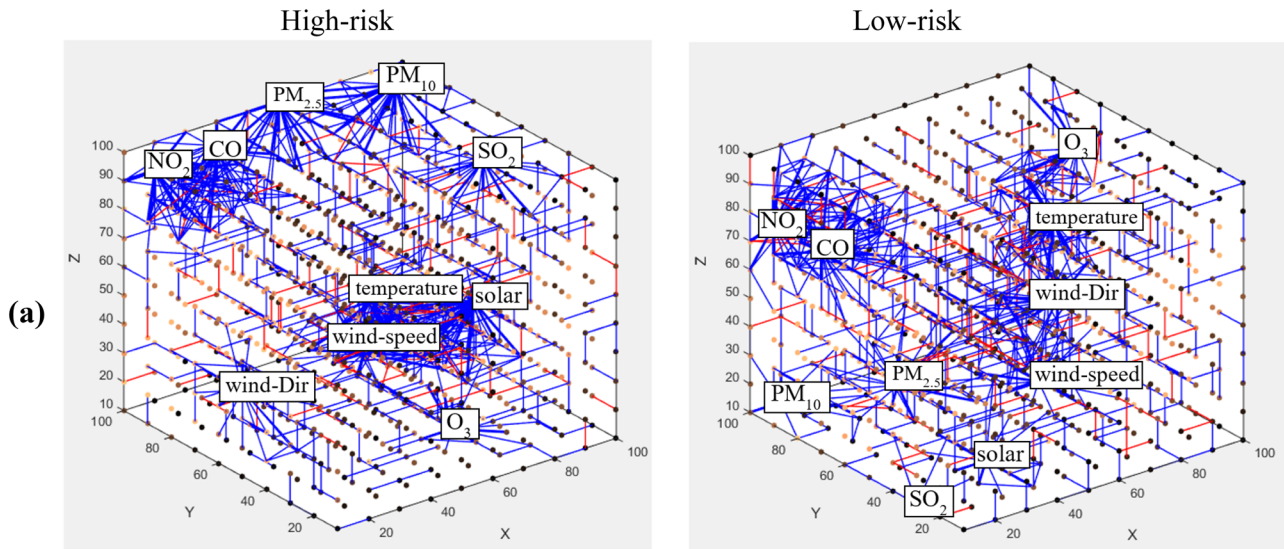
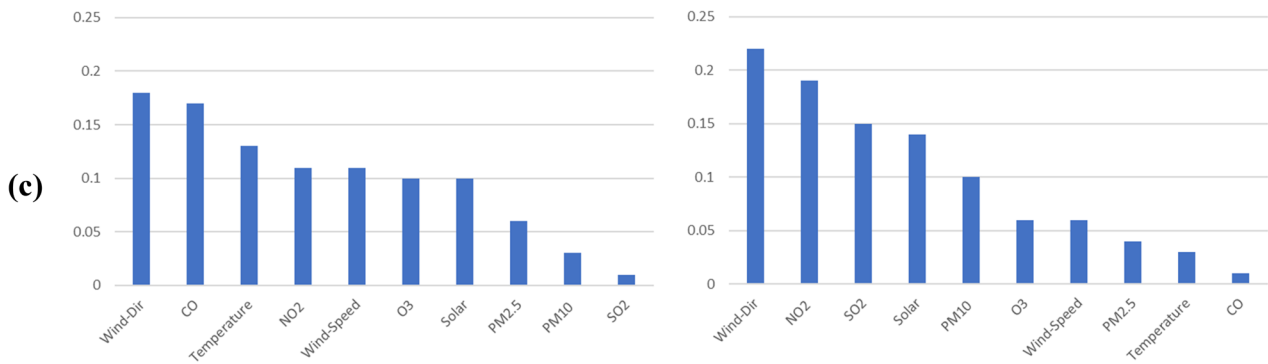Percentage of the activated neurons around environmental variables in an PSNN

(c)

◀**Fig. 8** (**a**) PSNN models were trained by 7-day environmental data in high-risk and low-risk periods for one randomly selected patient (21-year-old (female) who had a stroke on 18 Nov 2011 in Auckland, NZ) and had the following conditions: epilepsy, head injury, migraine and family history of heart attack, hypertension, diabetes. (**b**) Feature interaction network (FIN) shows the level of interactions between environmental features during the 7 days. (**c**) Percentage of the activated neurons in PSNN models presenting environmental variables is indicating the importance of these variables for stroke prediction within the cluster of patients closer to the selected individual

incidence and elevated NO_2, SO_2, and O_3 [6, 7]. A study in the USA discovered the relationships between stroke prevalence and exposure of $PM_{2.5}$ and O_3, advocating that further investigation on the association of pollution and stroke is vital [8].

Although the aforesaid studies have investigated a link between stroke occurrence and some environmental factors, the relationship between personal, clinical health variables, and certain environmental changes over time is not yet well
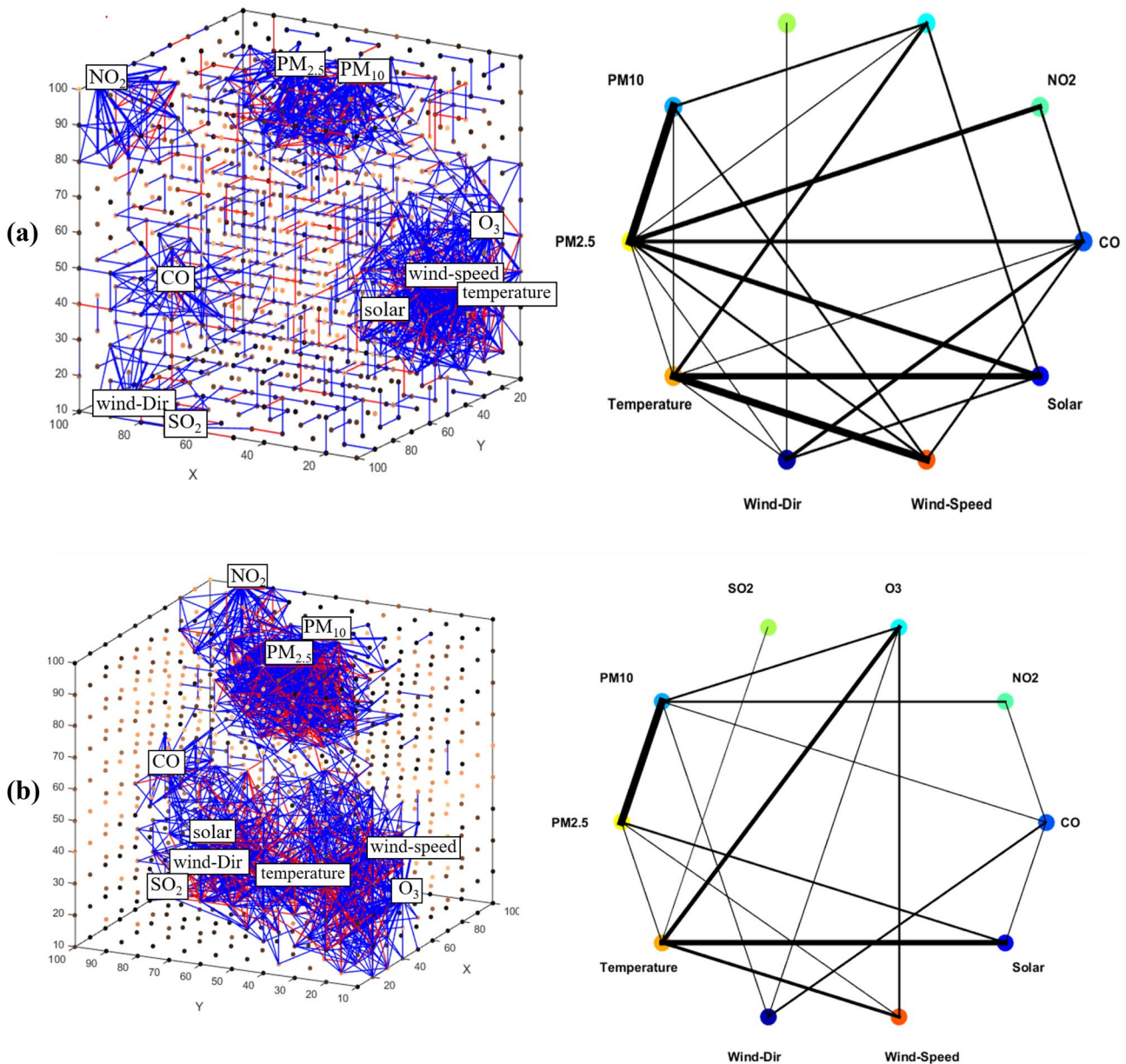
Fig. 9 Personalized profiling of two patients who had a stroke on (**a**) 29/Apr/2011 and (**b**) 30 Jan 2012 in Auckland, NZ, belonging to two clusters of subjects with the following information: age > 70, family history of stroke, high cholesterol, diabetes, vascular/heart disease; (left) PSNN connectivity trained with high-risk environmental data (encoded spikes from 7-day data). (Right) Feature interaction network shows the interactions between environmental features over 7 days, where the nodes represent the features, and the thickness of the lines shows the amount of information exchanged between them over time

investigated. The current study is an advancement on the existing predictive models of stroke by combining different data modalities for modeling complex interactions of risk factors. The personalized profiles of patients improved the models' interpretability so that an end-user (e.g., a medical practitioner) can comprehend what interactions between the environmental features have mostly increased the risk of stroke for an individual. It depicts a new avenue for practical implications of these findings and clinical use if the proposed algorithm will be fully tested, proved its robustness and accuracy, linked with the actual weather forecast, and shared as a usable device (e.g., a mobile app) with clinicians and family members of people with a higher risk of stroke for personalized prediction of stroke events. It will facilitate discussions with those at higher personalized risk of developing stroke within the next 7 days while they still retain the capacity to reduce the risk, regarding undertaking certain protective measures, such as escaping from a region where the determined environmental changes provoke stroke occurrence and moving closer to medical facilities, which would allow patients and families to receive medical care at an earlier stage in the disease process, and leading to improved prognosis and decreased morbidity and mortality.

Conclusion

The proposed personalized method and system allow for modeling and discovery of the relationship between personal health variables and environmental changes over several days (7 days) to estimate a probable risk of stroke. This system is built upon a cognitive-based computational architecture of spiking neural networks constituted of several methods in a pipeline that includes clustering of patients according to their personal data; developing personalized models of environmental time-series prior to the day of predicted risk of stroke event; classifying and predicting the high-risk environmental period; 3D visualization of models; and interpretation and knowledge discovery at an individual and a cluster-based approach. The personalized modeling approach and the developed machine learning algorithms can be used on other data, related to different populations, environmental, and clinical variables. In principle, the method can be used and tested on other time-windows of environmental data rather than the 7-day period used here as an example, to check if changes of environmental and other factors in any other timeframe can serve as risk factors for stroke.

Future work will include extracting spatio-temporal symbolic rules that represent the discovered associations between clinical and environmental variables for groups of individuals at high risk [23–25].

Declarations

References

1. Krishnamurthi RV, Ikeda T, Feigin VL. Global regional and country-specific burden of ischaemic stroke, intracerebral haemorrhage and subarachnoid haemorrhage: a systematic analysis

of the global burden of disease study 2017. Neuroepidemiology. 2020;54(2):171–9.

2. Johnson CO, et al. Global regional and national burden of stroke 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 2019;18(5):439–58.

3. Hankey GJ. Ischaemic stroke - prevention is better than cure (in English). J R Coll Phys Edinb. 2010;40(1):56–63. https://doi.org/10.4997/JRCPE.2010.111.

4. Feigin VL, et al. Global burden of stroke and risk factors in 188 countries during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet Neurol. 2016;15(9):913–24.

5. O'Donnell MJ, et al. Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study. Lancet. 2016;388(10046):761–75.

6. Tsai S-S, Goggins WB, Chiu H-F, Yang C-Y. Evidence for an association between air pollution and daily stroke admissions in Kaohsiung Taiwan. Stroke. 2003;34(11):2612–6.

7. Qian Y, et al. Epidemiological evidence on association between ambient air pollution and stroke mortality. J Epidemiol Commun Health. 2013;67(8):635–40.

8. Lisabeth LD, et al. Ambient air pollution and risk for ischemic stroke and transient ischemic attack. Ann Neurol. 2008;64(1):53–9.

9. Bokonjić R, Zec N. Strokes and the weather: a quantitative statistical study. J Neurol Sci. 1968;6(3):483–91.

10. Gordon PC. The epidemiology of cerebral vascular disease in Canada: an analysis of mortality data. Can Med Assoc J. 1966;95(20):1004.

11. Takahashi E, Sasaki N, Takeda J, Itō H. The geographic distribution of cerebral hemorrhage and hypertension in Japan. Hum Biol. 1957;29(2):139–66.

12. Alter M, Christoferson L, Resch J, Myers G, Ford J. Cerebrovascular disease: Frequency and population selectivity in an upper midwestern community. Stroke. 1970;1(6):454–65.

13. Feigin VL, Wiebers DO. Environmental factors and stroke: a selective review. J Stroke Cerebrovasc Dis. 1997;6(3):108–13.

14. Rogot E, Padgett SJ. Associations of coronary and stroke mortality with temperature and snowfall in selected areas of the United States 1962–1966. Am J Epidemiol. 1976;103(6):565–75.

15. Bull G, Morton J. Environment temperature and death rates. Age Ageing. 1978;7(4):210–24.

16. Wellenius GA, Schwartz J, Mittleman MA. Air pollution and hospital admissions for ischemic and hemorrhagic stroke among medicare beneficiaries. Stroke. 2005;36(12):2549–53.

17. Wordley J, Walters S, Ayres JG. Short term variations in hospital admissions and mortality and particulate air pollution. Occup Environ Med. 1997;54(2):108–16.

18. Shinkawa A, Ueda K, Hasuo Y, Kiyohara Y, Fujishima M. Seasonal variation in stroke incidence in Hisayama Japan. Stroke. 1990;21(9):1262–7.

19. Zhang Z-F, Yu S-Z, Zhou G-D. Indoor air pollution of coal fumes as a risk factor of stroke Shanghai. Am J Public Health. 1988;78(8):975–7.

20. Song S, Miller KD, Abbott LF. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. Nat Neurosci. 2000;3(9):919–26.

21. Kasabov N, Dhoble K, Nuntalid N, Indiveri G. Dynamic evolving spiking neural networks for on-line spatio-and spectrotemporal pattern recognition. Neural Netw. 2013;41:188–201.

22. Kasabov NK. Time-space spiking neural networks and brain-inspired artificial intelligence. Berlin: Springer; 2019.

23. Kasabov NK. NeuCube: a spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data. Neural Netw. 2014;52:62–76.

24. Kasabov NK, Hou Z-G, Feigin V, Chen Y. Method and system for predicting outcomes based on spatio/spectro-temporal data. In: Google Patents; 2020.

25. Kumarasinghe K, Kasabov N, Taylor D. Deep learning and deep knowledge representation in Spiking Neural Networks for Brain-Computer Interfaces. Neural Netw. 2020;121:169–85.

26. Doborjeh M, Doborjeh Z, Kasabov N, Barati M, Wang GY. Deep learning of explainable EEG patterns as dynamic spatiotemporal clusters and rules in a brain-inspired spiking neural network. Sensors. 2021;21(14):4900.

27. Othman M, Improved predictive personalized modelling with the use of Spiking Neural Network system and a case study on stroke occurrences data. In, et al. international joint conference on neural networks (IJCNN). IEEE. 2014;2014:3197–204.

28. Kasabov N, et al. Evolving spiking neural networks for personalised modelling classification and prediction of spatiotemporal patterns with a case study on stroke. Neurocomputing. 2014;134:269–79.

29. Kasabov NK. Evolving connectionist systems: the knowledge engineering approach. Berlin: Springer Science & Business Media; 2007.

30. Thorpe S, Gautrais J. Rank order coding. In: Bower JM, editor. Computational neuroscience. Boston: Springer; 1998. p. 113–8.

31. Verstraeten D, Schrauwen B, D'Haene M, Stroobandt D. An experimental unification of reservoir computing methods. Neural Netw. 2007;20(3):391–403.

32. Masquelier T, Guyonneau R, Thorpe S. Competitive STDP-based spike pattern learning. Neural Comput. 2009;21(5):1259–76.

33. Petro B, Kasabov N, Kiss RM. Selection and optimization of temporal spike encoding methods for spiking neural networks. IEEE Trans Neural Netw Learn Syst. 2019;31(2):358–70.

34. Kasabov N, Zhou L, Doborjeh MG, Gholami Z, Yang J. New algorithms for encoding learning and classification of fMRI data in a spiking neural network architecture: a case on modelling and understanding of dynamic cognitive processes. IEEE Trans Cogn Develop Syst. 2016;9(4):293–303.

35. Dhoble K, Nuntalid N, Indiveri G, Kasabov N. Online spatiotemporal pattern recognition with evolving spiking neural networks utilising address event representation, rank order, and temporal spike learning. In: IEEE World Congress on Computational Intelligence. Brisbane, Australia; 2012. p. 1–7.

36. Petro B, Kasabov N, Kiss RM. Selection and optimization of temporal spike encoding methods for spiking neural networks. IEEE Trans Neural Netw Learn Syst. 2019;31(2):358–70.

37. Doborjeh MG, Kasabov N, Doborjeh ZG. Evolving dynamic clustering of spatio/spectro-temporal data in 3D spiking neural network models and a case study on EEG data. Evol Syst. 2018;9(3):195–211.

38. Doborjeh MG, Wang GY, Kasabov NK, Kydd R, Russell B. A spiking neural network methodology and system for learning and comparative analysis of EEG data from healthy versus addiction treated versus addiction not treated subjects. IEEE Trans Biomed Eng. 2015;63(9):1830–41.

39. Doborjeh Z, et al. Spiking neural network modelling approach reveals how mindfulness training rewires the brain. Sci Rep. 2019;9(1):1–15.

40. Doborjeh ZG, Doborjeh M, Kasabov N. EEG pattern recognition using brain-inspired spiking neural networks for modelling human decision processes. In: 2018 International Joint Conference on Neural Networks (IJCNN). IEEE; 2018. p. 1–7.

41. Doborjeh ZG, Doborjeh MG, Kasabov N. Efficient recognition of attentional bias using EEG data and the NeuCube evolving spatio-temporal data machine. In: International Conference on Neural Information Processing. Springer; 2016. p. 645–53.

42. Doborjeh ZG, Doborjeh MG, Kasabov N. Attentional bias pattern recognition in spiking neural networks from spatio-temporal EEG data. Cognit Comput. 2018;10(1):35–48.

43. Doborjeh ZG, Kasabov N, Doborjeh MG, Sumich A. Modelling peri-perceptual brain processes in a deep learning spiking neural network architecture. Sci Rep. 2018;8(1):1–13.

44. Kasabov NK, Doborjeh MG, Doborjeh ZG. Mapping learning visualization classification and understanding of fMRI data in the NeuCube evolving spatiotemporal data machine of spiking neural networks. IEEE Trans Neural Netw Learn Syst. 2016;28(4):887–99.

45. Tu E, Neucube (st) for spatio-temporal data predictive modelling with a case study on ecological data. In, et al. international joint conference on neural networks (IJCNN). IEEE. 2014;2014:638–45.

46. Tu E, Kasabov N, Yang J. Mapping temporal variables into the neucube for improved pattern recognition predictive modeling and understanding of stream data. IEEE Trans Neural Netw Learn Syst. 2016;28(6):1305–17.

47. Kasabov N, et al. Evolving spatio-temporal data machines based on the NeuCube neuromorphic framework: design methodology and selected applications. Neural Netw. 2016;78:1–14.

48. Xing Y-F, Xu Y-H, Shi M-H, Lian Y-X. The impact of PM2. 5 on the human respiratory system. J Thorac Dis. 2016;8(1):E69.

49. O'Donnell MJ, Fang J, Mittleman MA, Kapral MK, Wellenius GA. Fine particulate air pollution (PM2. 5) and the risk of acute ischemic stroke. Epidemiology. 2011;22(3):422.

50. Santibañez DA, Ibarra S, Matus P, Seguel R. A five-year study of particulate matter (PM2. 5) and cerebrovascular diseases. Environ Pollut. 2013;181:1–6.

51. Lin H, et al. Ambient PM2. 5 and stroke: effect modifiers and population attributable risk in six low-and middle-income countries. Stroke. 2017;48(5):1191–7.

52. Wellenius GA, et al. Ambient air pollution and the risk of acute ischemic stroke. Arch Intern Med. 2012;172(3):229–34.

53. Eliasson B, Kogelschatz U. Ozone generation with narrow–band UV radiation. Ozone Sci Eng. 1991;13(3):365–73.

54. Brook RD, Brook JR, Urch B, Vincent R, Rajagopalan S, Silverman F. Inhalation of fine particulate air pollution and ozone causes acute arterial vasoconstriction in healthy adults. Circulation. 2002;105(13):1534–6.

55. Brook R, et al. Expert Panel on Population and Prevention Science of the American Heart Association Air pollution and cardiovascular disease: a statement for healthcare professionals from the Expert Panel on Population and Prevention Science of the American Heart Association. Circulation. 2004;109(21):2655–71.

56. Henrotin J-B, Besancenot J-P, Bejot Y, Giroud M. Short-term effects of ozone air pollution on ischaemic stroke occurrence: a case-crossover analysis from a 10-year population-based study in Dijon France. Occup Environ Med. 2007;64(7):439–45.

57. Montresor-López JA, et al. Short-term exposure to ambient ozone and stroke hospital admission: a case-crossover analysis. J Expo Sci Environ Epidemiol. 2016;26(2):162–6.

58. Feigin VL, et al. Global and regional burden of stroke during 1990–2010: findings from the Global Burden of Disease Study 2010. Lancet. 2014;383(9913):245–55.

59. Feigin VL, Norrving B, George MG, Foltz JL, Roth GA, Mensah GA. Prevention of stroke: a strategic global imperative. Nat Rev Neurol. 2016;12(9):501.