

## Article

# Driving Behaviour Style Study with a Hybrid Deep Learning Framework Based on GPS Data

Jingqiu Guo <sup>1</sup>, Yangzexi Liu <sup>1</sup>, Lanfang Zhang <sup>1</sup> and Yibing Wang <sup>2,\*</sup>

<sup>1</sup> Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 200092, China; 16522@tongji.edu.cn (J.G.); 18508230695@163.com (Y.L.); zlf2276@tongji.edu.cn (L.Z.)

<sup>2</sup> Institute of Intelligent Transportation Systems, Zhejiang University, Hangzhou 310058, China

\* Correspondence: wangyibing@zju.edu.cn; Tel.: +86-15355089088

Received: 29 March 2018; Accepted: 13 June 2018; Published: 6 July 2018



**Abstract:** Innovative technologies and traffic data sources provide great potential to extend advanced strategies and methods in travel behaviour research. Considering the increasing availability of real-time vehicle trajectory data and stimulated by the advances in the modelling and analysis of big data, this paper developed a hybrid unsupervised deep learning model to study driving behaviour and risk patterns. The approach combines Autoencoder and Self-organized Maps (AESOM), to extract latent features and classify driving behaviour. The specialized neural networks are applied to data from 4032 observations collected from Global Positioning System (GPS) sensors in Shenzhen, China. In two case studies, improper vehicle lateral position maintenance, speeding and inconsistent or excessive acceleration and deceleration have been identified. The experiments have shown that back propagation through multi-layer autoencoders is effective for non-linear and multi-modal dimensionality reduction, giving low reconstruction errors from big GPS datasets.

**Keywords:** driving behaviour analysis; deep learning; GPS data; risk pattern; AESOM

## 1. Introduction

Road accidents impose serious problems on society in terms of human, economic, medical and environmental costs. As the World Health Organization (WHO) announced, the total number of road fatalities was approximately 1.25 million in 2016 [1]. To understand the various factors associated with fatal and non-fatal road accidents is very crucial [2,3]. Intensive efforts have been made to understand human driving styles and the classification of drivers' risk patterns [4,5]. For example, the relationship between the sensitivity of the driver to complex driving situations and the vehicle control has been acknowledged as a major contributing factor in accidents [6]. Driving patterns and their influence on environment and fuel-use were also well studied [7]. In industry, automotive insurance companies integrate pay-as-you-drive or pay-how-you-drive modes for pricing. Based on driving aggressiveness, prices can be adapted to the individuals [8]. Moreover, characterizing driving behaviour can be particularly helpful for the development of vehicle automation [9,10].

Many studies on road safety have primarily focused on investigating factors associated with fatal and serious injuries and so less effort has been made on slight injuries or pre-crash scenarios mainly due to underreporting [3]. This fact may lead to biased conclusions in injury control and safety management. Although this topic has attracted great attention in the past, there is still much to be investigated, for example, the dimensions of driving patterns and their potential influence on road safety.

Traditionally, driving data are collected via travel surveys, simulator-based experiments or naturalistic driving studies. These contribute to the understanding of correlations between individual

demographics, road and traffic conditions, as well as safety. However, these experiments are costly and time consuming [11]. Innovative technologies and traffic data sources provide great potential to extend advanced strategies and methods in travel behaviour research. For instance, using smartphone sensors, we have an opportunity to incorporate data from classic techniques with data extracted from GPS (Global Positioning System), camera and accelerometer gyroscope [12]. Compared with other automobile sensors, such as OBD (On-Board Diagnostic), CAN (Controller Area Network) buses and cameras, GPS sensor data are often easier to collect, making them popular in large-scale research [13]. Typically, driving behaviour can be measured as multi-dimensional time-series data. In this paper, we focus on studying vehicle movement operations including speed change, acceleration/deceleration, turning, and their temporal combinations derived from GPS sensor data in a short and regular time interval (e.g., 1 s). Also, GPS data have plentiful characteristics in both time and space domains.

The main objectives of this study are:

- to develop the method which can extract highly correlated low-dimensional features in massive sensor data,
- the proposed hybrid method can also detect and reduce the negative effect of data noise and defects,
- based on the unsupervised pre-training, then to classify and cluster various driving patterns associated with higher road risk.

Deep learning has been recognized as the representative advance in the new statistical and computational paradigm for the new data era [14]. Data requirements are no longer an obstacle with the big GPS datasets, but the difficulties in network design and its interpretation are still the common bottlenecks. So far, applications of deep learning to large scale GPS data are limited. Motivated by the success of deep neural networks and considering the time and space characteristics of GPS data, we propose an unsupervised deep learning architecture to learn drivers' behaviour patterns from GPS data. The approach consists of two model components: feature learning by especially designed Autoencoder networks and feature clustering by SOM (Self-Organized Mapping) networks. The developed networks have been applied to a real large-scale GPS dataset to provide insight into driver behaviour and potential impacts in Shenzhen, South China.

The remainder of this paper is organized as follows. Section 2 reviews related works in the literature, with an emphasis on comparing statistical methods and neural networks in driving behaviour analysis. Section 3 presents the proposed deep learning approach. Then experimental studies on a large-scale GPS dataset in Shenzhen are discussed in Section 4. Section 5 presents a discussion and concludes the paper.

## 2. Related Works

### 2.1. Global Positioning System Data in a Travel Survey

In terms of data contents, driving simulations and naturalistic driving studies are the main sources on behaviour. Simulators allow for repetitive experiments and well-defined scenarios. Naturalistic driving studies, undertaken in natural conditions (no interference, no appearance of administrators and during daily driving) [15,16] provide the opportunity to observe the actual driving process with an unobtrusive high-precision data acquisition system. However, both approaches are costly and require a long-time frame. Novel vehicle technologies and the increasing degree of vehicle automation are changing driving patterns relatively quickly. Meantime, drivers generate continuous real-time vehicle trajectory data while using smartphone applications.

Big data era offers great opportunities to discover latent behaviour heterogeneities that are impossible with small samples. On the other hand, the huge sample size and high dimensionality bring statistical and computational challenges. For example, the identification of travel behaviour characteristics by means of GPS-based data has received intensive research effort in the last decade [17].

Compared with work on the social aspects of transportation which has been typically historical and off-line, new sensor technologies offer real-time computing and embedded applications with interactive big data for transportation studies. Zheng et al. [18] have suggested that the new social transportation field should focus on traffic analytics with big data using data mining, machine learning, and crowdsourcing mechanisms. Smartphone-based travel surveys are generally conducted using personal devices and navigation apps; they offer a key benefit in reducing both the cost of data collection and that of distributing and retrieving the hardware [19–21]. Most navigation applications utilize the mobile device, built-in GPS, to provide real-time location, route, traffic, parking, energy consumption and ride-sharing information [22].

The main advantages of GPS sensor data are: (i) unobtrusive data collection; (ii) large-scale of sample, e.g., millions of drivers; (iii) real-time and continuous dataset. Meanwhile, GPS sensors typically suffer two problems: high level of noise compared with dedicated instruments used in simulators and naturalistic driving studies and heterogeneous data sources. Some studies on data fusing have been conducted to integrate GPS data, accelerometer, gyroscope measurements for denoising and filtering [23]. It is also common to combine smartphone data with the data from CAN via the on-board diagnostic [24]. However, it is challenging to have a robust and generic approach to characterize behavioral patterns from GPS data.

## 2.2. Statistical Methods Versus Neural Networks in Driving Behaviour Analysis

In this section, two most popular approaches in driving behaviour analysis are compared: classic statistical algorithms e.g., [25,26] and computational intelligence e.g., [27]. Statistics covers collecting, organizing and interpreting data, as well as the mathematical analysis required to establish the statistical properties, distributions and parameters, and also to mine the spatio-temporal pattern from a probabilistic perspective [28]. Principal Component Analysis (PCA) has been identified as a successful method to learn essential features with the assumption that the input is independent as a Gaussian distribution [29]. Factor analysis is another widely used inter-dependency technique when the relevant set of variables shows a systematic inter-dependence and the objective is to search the underlying factors that create a commonality. For the results to be useful, the factors must be interpretable [7]. By using a nonlinear kernel function, Kernel Principal Component Analysis (KPCA) showed a higher accuracy rate than PCA and factor analysis in driving behaviour analysis [30]. However, when dealing with highly nonlinear data, computational intelligence may be considered more generic, accurate and intelligent [31].

Computational intelligence approaches combine learning, adaption, evolution, and fuzzy logic [6,32]. Neural networks with deep architectures are a major approach and have been widely applied in travel studies. Deep learning has been systematically compared to classical statistical models in comparative studies of classification and clustering, function approximation, and prediction problems [14,33]. For example, the Recurrent Neural Networks (RNN) is a dynamic classification algorithm used for recognizing time-series patterns in several domains [34]. One of the most successful deep learning neural networks to model big data in the space domain can be the Convolutional Neural Networks (CNN) which uses filters to find relationships between neighboring inputs [35]. In general, the literature suggested that Neural Networks (NN) are better performed when the idealized assumptions of statistical models (e.g., linearity, normality, Independent and identically distributed (IID)) are not valid, or the results from the statistical analysis are monotonous and difficult to interpret. Typically, multi-dimensional and complicated GPS data require more computational, flexible and nonlinear models. Meanwhile, massive dataset often contains more noise, defects and outliers, and NN are suitable in these cases.

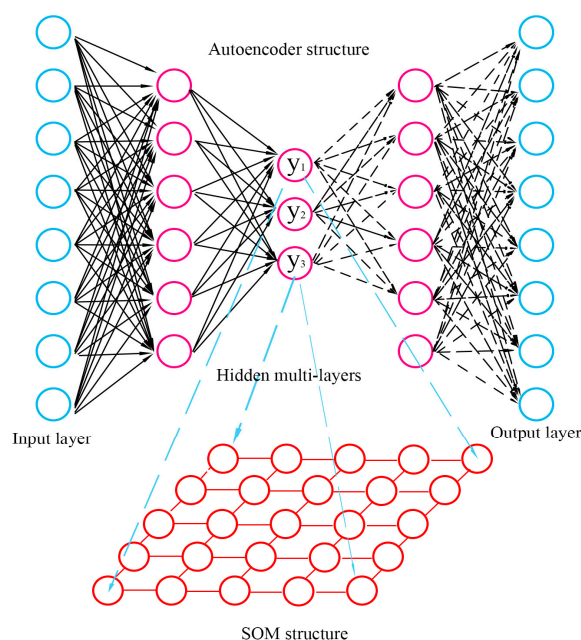
So far, applications of deep learning approaches to GPS-based data analysis are limited. Recently, deep multi-layer neural networks have been adopted in traffic prediction based on GPS data. Existing approaches typically follow the supervised learning algorithms, where the inputs are features and the labels are drivers' identification [36,37]. However, with a small training set and

diverse driver behavior, the learnt model may not work well. Furthermore, with a large number of drivers (e.g., more than 5000 candidates), the classification can be much more challenging than simply differentiating safe and unsafe drivers. Networks for three types of learning have been employed: supervised, unsupervised and a combination of the two. Although there are plenty of NN works, few attempts have been conducted on unsupervised learning for driving style identification and classification.

Unsupervised learning paradigms such as Autoencoder/Stacked Autoencoder, Restricted Boltzmann machine, and Long Short-Term Memory network are able to identify the performance features. For example, based on in-vehicle telematics, Jasper et al. [8] identified behavioural change among drivers following or during specific incentive using Long Short-Term Memory network. Dong et al. [38] proposed an autoencoder regularized deep neural network combining supervised and unsupervised learning. Ferrer and Ruiz [20] were using data extracted from GPS and GIS, compared five classification models including Decision Tree, Bayesian Network, Random Forest, Naïve Bayesian and Neural Network to identify travel patterns. In recent years, there is a growing body of evidence to suggest that competitive learning outperforms traditional clustering methods. In unsupervised and self-organizing neural networks, the two dominant models are the SOM and adaptive resonance theory (ART), both of which are based on competitive learning [39]. Conclusively both statistical algorithms and neural networks show advantages and limitations in transportation behavioural research. Despite their differences, there is a growing trend to combine statistical algorithms and neural networks into one platform [27], especially in the fields of causality investigation, the analysis of big data and model development and evaluation.

### 3. Methodology

The aims of this research were to (i) develop an effective approach that can extract the low-dimensional high-level features of driving behavior; and (ii) accurately explore the hidden behaviour sub-groups across a heterogeneous population. In light of the shortcomings of existing methods, we propose a deep learning framework to study driving characteristics on GPS data, in an unsupervised feature learning and classification architecture, which is called AESOM (Autoencoder-Self Organizing Mapping), as shown in Figure 1.



**Figure 1.** Overall architecture of Autoencoder-Self Organizing Mapping (AESOM); SOM: self-organizing map.

For simplicity, we consider GPS data as the unique raw input source. Nevertheless, this framework can be generalized to work with other types of sensor dataset and rich driving contexts. We first introduce the autoencoder networks that read GPS data as inputs, and learn and extract low-dimensional driving behaviour features. Then we discuss classification and clustering of high-level driving features by using the SOM networks.

### 3.1. Selection of Input Features

Previous research has extensively studied the classification process, the input data analysis and the algorithms used to predict and label drivers into specific driving styles [40]. Therefore, it is hypothesized that actions of a driver in a specific category of driving can be represented and predicted due to the definitive and measurable nature of driving styles [11], such as aggressive drivers, conservative or slow drivers, inattentive drivers, drunk drivers. However, to detect unsafe drivers, for most learning and classification techniques, a pre-set ‘normal’ driving profile has to be defined as reference, typically, using a discrete scale with several levels [41]. As opposed to human-labelled learning, our feature extraction approach is unsupervised without any prior knowledge and does not rely on a prior definition. The objective is first to develop an adaptive, multilayer “encoder” network to transform the higher-dimensional raw representation into a low-dimensional code and a similar “decoder” network to promise the minimal loss.

In general, an aggressive or inattentive driving pattern has been associated with risky speeding profiles, inconsistent or excessive acceleration/deceleration, and improper vehicle position maintenance. For simplicity, we focus on features of speed and acceleration only in this part. The GPS data also includes vehicle orientation information (by degree) per second. The feature transformation related to turning behaviour can be modelled and clustered following the same steps. The details are discussed in Section 4.3.

#### 3.1.1. Parameters Related to Speed

The effect of speed on road safety has gained considerable attention in literature. Parameters such as speed limits, mean speed and the speed variance have been examined in order to define the underlying relationships [2]. For instance, in China metropolitan areas, the speed limit is often no more than 80 km/h. Thus, for a maximum speed limit  $v_{max}$ , we define  $v_f$  as the threshold speed if there is an over speed tendency when the vehicle runs at a relatively high speed. The scale of time duration when its speed is over  $v_f$  is calculated in Equation (1),

$$\delta = \frac{T_f}{T_c} \times 100\%, \quad (1)$$

where  $T_c$  is the total travelling time of this vehicle on road;  $T_f$  is the duration of time when speed exceeds  $v_f$ . Besides, the mean and the unbiased estimation of standard deviation of the vehicle’s speed are included due to the clear correlation between road accidents and a wide variation in speed.

$$v_a = \frac{1}{n} \sum_{m=1}^n v_m, \quad (2)$$

$$v_s = \sqrt{\frac{1}{n-1} \sum_{m=1}^n (v_m - v_a)^2}, \quad (3)$$

where  $v_m$  is the instantaneous speed of the vehicle collected by GPS sensor at the time  $m$ ;  $n$  is the sample size.

### 3.1.2. Parameters Related to Acceleration

Acceleration and jerk variance evaluation for vehicles are often employed to detect anomalous behaviour in driving risk analysis. Thus, multi-dimensional acceleration-related parameters need to be incorporated into the model, which are the unbiased estimation of standard deviation  $a_s$ , positive (negative) standard deviation  $a_s^+(a_s^-)$ , positive (negative) average  $a_m^+(a_m^-)$  of acceleration.

$$a_a = \frac{1}{n} \sum_{m=1}^n a_m, \quad (4)$$

$$a_s = \sqrt{\frac{1}{n-1} \sum_{m=1}^n (a_m - a_a)^2}, \quad (5)$$

$$a_m^\pm = \frac{v_m - v_{m-1}}{t}, \quad (6)$$

$$a_a^\pm = \frac{1}{n} \sum_{m=1}^n a_m^\pm, \quad (7)$$

$$a_s^\pm = \sqrt{\frac{1}{n-1} \sum_{m=1}^n (a_m^\pm - a_a^\pm)^2}, \quad (8)$$

where  $a_m$  is the instantaneous acceleration of the vehicle collected by GPS at the time  $m$ ;  $a_a$  is the average acceleration of the vehicles in the dataset.

Thus, on an experimental basis, we first selected eight features transformed from sequences of raw GPS data to serve as the behavior-to-vec  $X$  of drivers, where  $X$  is constructed as follows:

$$X = [\delta, v_a, v_s, a_s, a_a^+, a_s^+, a_a^-, a_s^-]$$

### 3.2. Learning Features with Autoencoder

The Autoencoder is a feedforward neural network with a symmetrical structure and an odd number of layers [42–44], by minimizing the reconstruction error between the input data at the encoding layer and its reconstruction at the decoding layer, in the mapping weight vectors. Autoencoder has a unique training process to transform the dimension of the data, e.g., from high level to lower one, or vice versa. It guarantees that every neuron in input/output layers can have a one-to-one correspondence to the feature. For the simplest case with a hidden layer smaller than the input/output layer and linear activations only, the autoencoder implements a compression scheme and performs equally as PCA [45]. Recent studies found that nonlinear autoencoders are capable of classifying certain types of multimodal and nonlinear domains accurately, and so reveal much deeper connections between variables [46].

To enhance the performance of an autoencoder, the input vector needs to be element-wise normalized to

$$x_i = \frac{x_i - \bar{x}_i}{\sigma_i} \quad (9)$$

where  $\bar{x}_i$  is the average value of each feature and  $\sigma_i$  is the variance. We take the normalized vector  $X$  as the input of the autoencoder and define  $V$  as the behaviour training set which is ready to feed into the autoencoder with a size of  $M$ .

$$V = \{X_1, X_2, \dots, X_n\}^T, n \in M$$

where  $n$  is the size of the dataset. The model is designed as a multiple layer structure ( $\geq 3$ ), aiming to compress and extract principal latent features for further analysis, such as classification and regression. The first layer is the input layer, the last one is the output layer, and others are hidden layers.

Autoencoders offer a way of defining a nonlinear form of  $h_{W,b}(x)$  with parameters  $W, b$  that can fit to the dataset. Moreover, the goal of iterating the autoencoder is to learn an identity function  $h_{W,b}(x) = \hat{x}_i \approx x_i$ , where  $\hat{x}_i$  is the output vector corresponding to the input vector  $x_i$ . We let  $n_l$  denote



the number of layers in our network, i.e.,  $n_l = 5$ ;  $W_{ij}^{(l)}$  denotes the parameter (or weight) associated with the connection between unit  $j$  in layer  $l$ , and unit  $i$  in layer  $l + 1$ ;  $b_i^{(l)}$  is the bias term associated with unit  $i$  in layer  $l + 1$ . The weights in every layer start with a random value and iterate when trying to minimize the loss between the original input and its output. We use tangent function  $\tanh(\cdot)$  as activation function between layers as  $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ .

Using the chain rule to back propagate error derivatives through deeper layers to shallower ones, the gradient method can easily manage the process. Taking the third layer as an example ( $l = 3$ ), with a sparse constraint on the hidden units, in other words for most of its elements to be zero, the average value  $\hat{p}_j$  of activation degree of hidden neuron  $j$  on the training set is:

$$\hat{p}_j = \frac{1}{M} \sum_{i=1}^M a_j^{(3)}(x^{(i)}) \quad (10)$$

where  $a_j^{(3)}$  denotes the activation degree of hidden neuron  $j$ . Parameter  $\rho$  specifies the desired level of sparsity whose value is close to zero (e.g., 0.01). Now we can define the overall cost function to be

$$J(W, b) = \frac{1}{M} \sum_{i=1}^M (\hat{x} - x)^2 + \beta \sum_{j=1}^s [\rho \cdot \log \frac{\rho}{\hat{p}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{p}_j}] \quad (11)$$

where  $s$  is the number of neurons in the hidden layer. As shown in Equation (11), we apply Kullback-Leibler divergence as the penalty term,  $\beta$  is the weight of the penalty term. In the BP (Back Propagation) training process, the update of weights is calculated as:

$$\varphi^{(3)} = \left[ \sum_{j=1}^s W_{ji}^{(3)} \varphi^{(4)} + \beta \left( \frac{\rho}{\hat{p}_j} + \frac{1 - \rho}{1 - \hat{p}_j} \right) \right] \cdot f'(z_i^{(3)}) \quad (12)$$

In this study, among the hidden layers, it is a kind of layer-wise-pre-training process, in other words, the shallow layers ( $l = 2$ ) are used to learn the fairly simple and straightforward regulations from units, and then the deeper hidden layers ( $l = 3$ ) are applied to learn latent regulations or multimodal domains.

### 3.3. Classification with Self-Organizing Map

Initially researchers employed SOM network for quantization of colour images. By adjusting a quality factor, the network is able experimentally to produce images of much greater quality than existing methods. In a refined version of the SOM, the output can be used for a controlled training of the next layer network, called as an unsupervised clustering method hierarchical SOM [47]. This method provides a natural measure for the distance of a neuron from a cluster by giving appropriate weights to all the neurons belonging to the cluster, and so produces clusters that match the desired classes better than the direct SOM or the classical k-means.

Classification and clustering of individual drivers is the primary area of interest in this driving behaviour study. SOM learn on their own through unsupervised competitive learning, where the neurons need to race for the opportunities of activation. This mechanism has the effect that only one neuron can be activated by the input stimuli at any time  $t$ . For classification, a major advantage of SOM is to be able to minimize the influence of the noise data and so outperform other classic clustering techniques e.g., the K-means method [48]. Furthermore, SOM has gained in popularity due to the ability to preserve the topology in projections, where the topology form is often represented as a rectangular  $I \times J$  grid.

Self-Organizing Maps are so named because no inspection is required and they assign themselves according to the weights given to the input nodes. Another unique aspect of SOM is termed Vector Quantization, which is a data compression technique of representing multi-dimensional data in

lower dimensional space (1-D or 2-D). To keep the same dimension as the input vector, Neuron  $n_{ij}$  is synchronized with a weight vector  $w_{ij}$ . Before the training process, weight vectors are initialized at random values generated from the uniform distribution ranging from 0 to 1. When  $n_{ij}$  receives the input vector  $y$  at every iteration  $t$ , the net input  $s_{ij}$  is calculated using Euclidean Distance in Equation (13), i.e., every input neuron is compared with the trained data based on collection frequency.

$$s_{ij}(t) = ||y(t) - w_{ij}(t)|| \quad (13)$$

The activity level  $a_{ij}$  is calculated at each iteration through the exponential function:

$$a_{ij}(t) = \exp\left(-\frac{s_{ij}(t)}{\sigma}\right) \quad (14)$$

where  $\sigma$  is the exponential factor. The  $\exp(\cdot)$  is to normalize and promote the difference between high and low degrees of activity of neurons. Thus, the winner neuron with strongest activation is selected as:

$$c = \underset{i,j}{\operatorname{argmax}} a_{ij}(t) \quad (15)$$

Then weights  $w_{ijk}$  are updated as below:

$$w_{ijk}(t+1) = w_{ijk}(t) + \alpha(t) \cdot G_{ijc}(t) [x_k(t) - w_{ijc}(t)] \quad (16)$$

where  $\alpha(t)$  is the learning rate ranging from 0 to 1, which controls how fast the state of SOM changes and how accurately the algorithm learns.  $G_{ijc}(t)$  is the neighborhood function which reflects the relationship between activated neuron and inactivated ones measured by the spatial distances as shown in Equation (17).

$$G_{ijc}(t) = \exp\left\{-\frac{r_c - r_{ij}}{2\sigma^2(t)}\right\} \quad (17)$$

where  $r_c, r_{ij} \in R^2$  are location vectors of neurons.

To obtain a form of input neurons for the SOM, we first transfer the encoded low-dimensional high-level parameters, into the learned feature extraction matrix. Then we transform them to be the input pattern of the SOM by combining with a 3-D vector  $\beta = (\beta_0, \beta_1, \beta_2)$ , and the competitive layer of SOM is set to be a practical 2-D mapping to obtain aggregated information. In the next part, we demonstrate the application of AESOM networks in a case study using a large-scale GPS dataset.

## 4. Experiments

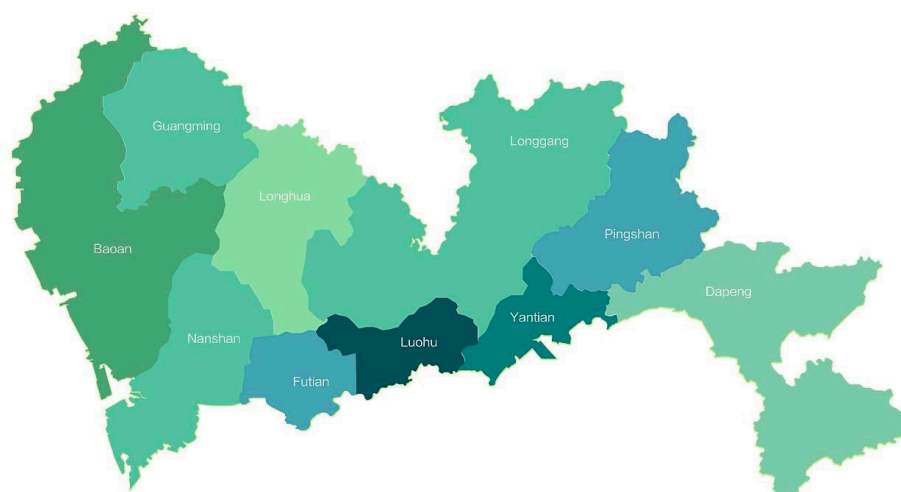
### 4.1. Datasets

In this case study, we use smartphone GPS data from ‘the City of Shenzhen Mapping’ database (source from Shenzhen Urban Transport Planning Center, Shenzhen, China) to understand the local driving behaviour in Shenzhen, China. The municipality of Shenzhen covers an area of 1991 square kilometers including urban and rural areas, with a total population of approximate 12 million. The city has an elongated shape measuring 81.4 km from east to west while the shortest section from north to south is only 10.8 km (Figure 2). Shenzhen was established as a Special Economic Zone in 1980, and so the road system is relatively modern and well planned.

As shown in Table 1, each subset corresponds to a unique and continuous period between January and June in 2017. The data in each set were collected in seven consecutive days. GPS are measured at every 1 s, resulting in a 3TB database. Given the amount of data, we have a large number of samples for almost every major road and expressway in Shenzhen. To review one subset in Table 1 as an example, there are approximately 2 billion points (each with unique latitude and longitude) collected in Set 5 from 1 May 2017 to 7 May 2017, from Monday to Sunday. The raw data may encounter both



user and system errors. The system errors are mainly due to technical issues such as signal reflection, phone battery, canyon effects and network connection that disturbs data transfer between the user and the server. To address the issue of noise and outlier, most GPS receivers employ proprietary filtering algorithms to compensate for data points beyond variances. Thus, the software embedded within the receiver automatically provides certain level of data correction. Second, additional measures of reliability can help identify questionable data, and numerous techniques can filter the data based on these measures (e.g., Pauta criterion). Third, one advantage of this proposed deep learning approach is to reduce/manage the negative effects of defects during feature extraction.



**Figure 2.** The city of Shenzhen (source from Shenzhen Urban Transport Planning Center).

**Table 1.** Description of the raw Global Positioning System database.

Set	Period	Total GPS Points (Millions)	Total Device ID (Thousands)
1	1–7 January 2017	1865	301
2	1–7 February 2017	1774	222
3	1–7 March 2017	2112	346
4	1–7 April 2017	2042	292
5	1–7 May 2017	1990	296
6	1–7 June 2017	2069	280

#### 4.2. Autoencoder-Self Organizing Mapping Network Application—Test One

To obtain samples that are most representative of the entire population, we followed a two-stage sampling process. First the whole dataset was divided into 42 subgroups by date (7 days in a month for 6 months). Then we selected the top 96 ids which contain most of the valid GPS points in each day. Finally, we obtained 4032 ids as the driving behaviour training set for the AESOM neural networks (Table 2). To verify both the feasibility and accuracy of the autoencoder, we performed some experiments and decide to employ the structure consisting of an encoder with one input layer and two hidden layers, neuron size as 8-6-3 respectively (Figure 1). This also determines the extraction feature matrix structure and a 3-D vector  $\beta = (\beta_0, \beta_1, \beta_2)$  to transform the significant components into input neurons of the SOM for clustering. The lattice of a competitive layer in SOM was set as a  $5 \times 5$  grid, with consideration of both computational cost and classification performance.

In this AESOM framework, the main objective of an autoencoder is to detect the structure of a large multivariate dataset (data patterns and relations) and to implement a compression scheme. In addition, it learns to what extent each component is associated with each input variable and how much the set of components explain the variability of the original dataset. After obtaining the component vector

$\hat{y}_n$  (the output of layer 3), we can understand and name each factor, observing the contribution of all the variables. Table 3 shows the output  $\hat{y}_n$ , where the loss value determined by  $J(W, b)$  is only 0.08, indicating a well-qualified neural network performance. The relationship between  $\hat{y}_n$  and input variables  $x_i$  is displayed in Table 4.

According to the feature extraction matrix in Table 4,  $\hat{y}_1$  and  $\hat{y}_3$  display strong relationships with acceleration/deceleration driving features ( $[a_s, a_a^+, a_s^+]$  and  $[a_s, a_a^-, a_s^-]$ ). Specifically,  $\hat{y}_1$  reflects acceleration with “+” sign while  $\hat{y}_3$  represents deceleration of vehicles (with “−” sign). In contrast,  $\hat{y}_2$  reflects speeding behaviour ( $[\delta, v_a, v_s]$ ).

**Table 2.** Parameter data of characteristics (portion).

Id	$\delta$	$v_a$	$v_s$	$a_s$	$a_a^+$	$a_s^+$	$a_a^-$	$a_s^-$
1	0.068	0.220	−0.083	−0.441	−0.755	−0.467	0.778	−0.330
2	−0.623	−0.112	−0.056	−0.847	−1.050	−0.771	1.070	−0.794
3	−0.381	−0.138	1.650	0.435	0.468	0.452	−0.418	0.355
4	−0.819	0.496	0.750	0.844	−0.135	1.101	0.179	0.847
5	−0.935	−1.777	−1.347	−1.056	−1.146	−0.931	1.154	−0.994
...	...	...	...	...	...	...	...	...
4031	0.057	0.315	−0.443	0.334	−0.405	0.543	0.392	0.557
4032	−0.804	0.159	−0.594	1.905	0.744	2.291	−0.727	2.323

**Table 3.** Encoded parameter matrix in the hidden layer  $l = 3$ .

Id	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$
1	2.07276	1.01655	1.50635
2	1.33207	0.99798	1.14364
3	5.74891	4.37044	5.20796
4	5.20190	1.46239	4.15118
5	1.20631	0.73165	1.06726
...	...	...	...
4031	4.10804	0.86285	2.75928
4032	8.79163	1.85217	5.82063

**Table 4.** Extraction of the most significant components.

Input Variable	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$
$\delta$	4.00863	<b>3.21656 *</b>	2.50600
$v_a$	4.15231	<b>2.42050 *</b>	2.60216
$v_s$	4.18947	<b>3.13631 *</b>	2.85290
$a_s$	<b>4.72042 *</b>	2.39948	<b>3.63525 *</b>
$a_a^+$	<b>5.49825 *</b>	2.13975	2.68946
$a_s^+$	<b>4.66887 *</b>	1.78941	3.17082
$a_a^-$	3.91298	2.13656	<b>3.47919 *</b>
$a_s^-$	4.09028	1.79880	<b>3.18694 *</b>

\* Bold and italic, high degree of association.

There is a growing body of evidence to suggest several road safety benefits are associated with reduced speed variability between vehicles. Specifically, increased speed variation may disturb homogenised traffic flow and increase the likelihood of conflict situations caused by human behaviour [49]. Considering the combination of acceleration and deceleration ( $\hat{y}_1$  and  $\hat{y}_3$ ), we set the coefficient  $\beta$  to be (1, 0, 1), and correspondingly design the SOM networks to cluster the 2-D inputs into 4 classes as shown in Table 5. It is good to find almost half driving in Shengzhen metro with consistent speed. Only 3.2% drivers show heavily variable speed, thus, we call this small group “Neurotic” drivers.

Based on  $\hat{y}_2$ , the SOM networks produced four distinct clusters as displayed in Table 6. 3.69% drivers would be classified as consistently speeding. This smallest group can be labeled as an “Aggressive” class. Even though a small percentage, there are over 3 million vehicles in Shenzhen and around 1.7 million vehicles on road each day in 2017. Thus, the actual volume of aggressive drivers on road daily in Shenzhen can be up to 70,000. The rest 53.08% (C2 + C3) show light to moderate risky speeding profiles.

It is noted that the outputs from the SOM networks based on  $\hat{y}_1$  or  $\hat{y}_3$  are in only three distinct clusters. As presented in Table 7, 66.34% of drivers (in C1) prefer to decelerate in a relatively smooth style. However, there are 6.48% of drivers who exhibit inconsistent or excessive accelerations (harsh take-off), labelled as “Inattentive” drivers. As expected, clustering based on  $\hat{y}_3$  indicates a similar distribution to that on  $\hat{y}_1$ , that about 93.99% of drivers (C1 + C2) constitute the norm, while 6.01% decelerate frequently, as shown in Table 8. They are more likely to closely tailgate and suddenly brake.

In conclusion, in Shenzhen, by and large drivers conformed to road rules, staying within the confines of the speed limit, with no harsh braking or sharp accelerating. Drivers in a small group were prone to acceleration as well as deceleration. This kind of motion can create high risk of accidents. The driving behaviour patterns have various physical, psychological and incidental aspects that are measurable. Driver behaviour is related not only to the driver’s character and socio-economics, but also to education, training, police enforcement, etc.

**Table 5.** Clusters of driving behaviour based on  $\hat{y}_1$  and  $\hat{y}_3$  ( $\beta = (1, 0, 1)$ ).

Class	Variable Speed Pattern	Percentage
C1	None	49.00%
C2	Slight	36.09%
C3	Moderate	10.81%
C4	Heavy	3.20%

**Table 6.** Clusters of driving behaviour based on  $\hat{y}_2$  ( $\beta = (0, 1, 0)$ ).

Class	Speeding Behavior	Percentage
C1	None	43.23%
C2	Slight	38.62%
C3	Moderate	14.46%
C4	Heavy	3.69%

**Table 7.** Clusters of driving behaviour based on  $\hat{y}_1$  ( $\beta = (1, 0, 0)$ ).

Class	Acceleration Behaviour	Percentage
C1	Slight	66.34%
C2	Moderate	27.18%
C3	Heavy	6.48%

**Table 8.** Clusters of driving behaviour based on  $\hat{y}_3$  ( $\beta = (0, 0, 1)$ ).

Class	Deceleration Behaviour	Percentage
C1	Slight	63.19%
C2	Moderate	30.80%
C3	Heavy	6.01%

#### 4.3. Autoencoder-Self Organizing Mapping Network Application—Test Two

In the second experiment, to investigate improper vehicle position maintenance, we added two vehicle lateral orientation features  $w_a$  and  $u_w$  from the raw sequences of GPS data. Here  $w_a$  is the

instantaneous angular velocity of the vehicle;  $u_w$  is the angular acceleration. Thus, a ten dimensional  $X$  vector was formulated as follows.

$$X = [\delta, v_a, v_s, a_s, a_a^+, a_s^+, a_a^-, a_s^-, w_a, u_w]$$

We keep the same autoencoder structure as in test one, consisting of an encoder with one input layer and two hidden layers, neuron size as 8-6-3 respectively. The new loss value is 0.03, indicating a better neural network performance compared with the network in test one. The results are presented in Tables 9–12.

**Table 9.** Feature extraction matrix in test II.

Features	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$
$\delta$	0.8360	0.8874	<b>1.9262 *</b>
$v_a$	0.9625	0.9234	<b>1.9001 *</b>
$v_s$	1.0015	<b>1.0148 *</b>	1.7624
$a_s$	0.9500	<b>1.1408 *</b>	1.8224
$a_a^+$	0.9428	<b>1.0362 *</b>	1.7620
$a_s^+$	1.0013	<b>1.0418 *</b>	1.8552
$a_a^-$	<b>1.0294 *</b>	0.8947	1.8071
$a_s^-$	<b>1.2016 *</b>	1.0012	1.8690
$w_a$	<b>1.0720 *</b>	0.8628	<b>2.0853 *</b>
$u_w$	<b>1.1305 *</b>	0.8509	<b>2.0830 *</b>

\* Bold and italic, high degree of association.

**Table 10.** Test II Clusters of driving behaviour based on  $\hat{y}_1$ .

Class	Turning and Deceleration	Percentage
C1	Slight	35.81%
C2	Moderate	40.67%
C3	Heavy	23.52%

**Table 11.** Test II Clusters of driving behaviour based on  $\hat{y}_2$ .

Class	Acceleration Behaviour	Percentage
C1	Slight	58.13%
C2	Moderate	33.93%
C3	Heavy	7.94%

**Table 12.** Test II Clusters of driving behaviour based on  $\hat{y}_3$ .

Class	Turning and Speeding	Percentage
C1	None	12.00%
C2	Slight	46.88%
C3	Moderate	39.83%
C4	Heavy	1.29%

Test two considers variations in the lateral and longitudinal position of the vehicle. According to the extraction matrix in Table 9,  $\hat{y}_1$  presents a strong correlation between angular velocity and vehicle deceleration features;  $\hat{y}_2$  reflects association with speed and acceleration with “+” sign behavior, while  $\hat{y}_3$  displays a strong relationship with the combination of lateral and longitudinal speed features.

In contrast to the clustering results in test one with a small class defying driving norms, the SOM networks in test two produced three distinct clusters based on  $\hat{y}_1$  and there are 23.52% of drivers in C3 who conduct sharp turning with deceleration (Table 10). Typical scenarios can be turning at

intersections, where they tend to turn the steering wheel suddenly with harsh braking simultaneously. Another differentiating factor of the way drivers turn is that we see more extreme lateral acceleration with high speed in a small class C4 (1.29%) in Table 12 based on  $\hat{y}_3$ , indicating a higher risk of accidents.

Improper vehicle lateral position maintenance and inconsistent or excessive angular acceleration/deceleration have been identified as major contributing behavioural characteristics that influence road safety. This proposed AESOM approach provides a good opportunity to combine feature learning and classification into an integrated deep learning framework to discover latent patterns and values from mega sensor data. The clustering results display the heterogeneous driving style profiles across the population. By adding the vehicle lateral orientation parameters to the neural networks, experiments verify the advantages of AESOM when dealing with high dimensionality.

## 5. Discussion and Conclusions

Sensors have made it both technically and economically feasible to review driving behaviour in natural surroundings on a large scale, through unobtrusive data collection and without experimental control. However, raw sensor data, that requires more than two or three dimensions to represent, can be difficult to interpret. One method of simplification is to assume that the data of interest lie on an embedded nonlinear manifold within the higher-dimensional space. Nonlinear dimensionality reduction algorithms have been convinced to perform much better than linear algorithms in the field of computer-vision. As an important branch of nonlinear dimensionality reduction, mapping methods can be viewed as a preliminary feature extraction step, after which pattern recognition algorithms are applied. Both autoencoders and self-organised mapping is prominent mapping learning algorithms. However, it is the first time we integrated these two nonlinear dimensionality reduction techniques into a hybrid unsupervised deep learning architecture, to learn and classify driving behaviour using GPS data.

In recently years, researchers have studies a series of model-driven deep learning methods, which showed their feasibilities and effectiveness in various fields. It can retain the powerful learning ability of the deep-learning approach, and may overcome the difficulty in network topology selection [50,51]. In this work, we made efforts to explore a model-driven deep learning framework that can balance the flexibility and appositeness in behaviour study. As an innovative architecture, some questions remain open and require further investigations. For example, can this model be improved by imposing certain form of sparsity on the representations they learnt? Popular dimensionality reduction approaches, whether linear ones like PCA, Independent component analysis and factor analysis, or non-linear such as Locally-Linear Embedding or Modified Locally-Linear Embedding, map each example to the same low-dimensional space. However, it has been discussed that in favor of sparsity, it would be practical and more efficient to map each example to a variable length representation.

Several potential directions are open for future exploration. First, by adding supervisory information into unsupervised feature learning to reconstruct our unsupervised AESOM, it may improve the quality of driving pattern representation. Second, further work is being conducted to study the performance of AESOM framework in detecting abnormal driving, and to refine the deep learning architecture to address the prediction of crash and near-crash events. Meanwhile, identifying behavioural change among drivers during or following specific events, time periods, or new transportation regulations, can be an important application for the proposed method.

The following conclusions were obtained:

1. Compared with the state-of-the-art modelling and analysis methods, the experiments have shown that back propagation through the multi-layer autoencoders were effective for non-linear and multi-modal dimensionality reduction, producing low reconstruction errors on big GPS datasets.
2. The driving behaviour features and clusters learned by the AESOM networks were fairly interpretable.

Most of traffic parameters were found to have mixed effects on road network. Thus, by extracting highly correlated time-series data of latent features and clustering into driving risk groups, this approach can be an effective tool for proactive road management strategies.

**Author Contributions:** The authors, J.G and Y.L. have all contributed to preparing this manuscript, L.Z. calibrated the results. The first two authors share equal contribution.

**Funding:** The research reported in this paper was supported in part by the Zhejiang Qianren Program (2013–2018), Shanghai Pujiang Program (17PJJC097), the Fundamental Research Funds for the Central Universities, and the National Natural Science Foundation of China (51478428; 71771200; 71671126).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. World Health Organization (WHO). *Road Traffic Injuries*; World Health Organization: Geneva, Switzerland, 2016.
2. Theofilatos, A.; Yannis, G. A review of the effect of traffic and weather characteristics on road safety. *Accid. Anal. Prev.* **2014**, *72*, 244–256. [[CrossRef](#)] [[PubMed](#)]
3. Potoglou, D.; Carlucci, F.; Cirà, A.; Restaino, M. Factors associated with urban non-fatal road-accident severity. *Int. J. Inj. Contr. Saf. Promot.* **2018**, 1–8. [[CrossRef](#)] [[PubMed](#)]
4. Mitrovic, D. Reliable method for driving events recognition. *IEEE Trans. Intell. Transp. Syst.* **2005**, *6*, 198–205. [[CrossRef](#)]
5. Sagberg, F.; Selpi, Piccinini, G.F.; Engström, J. A Review of Research on Driving Styles and Road Safety. *Hum. Factors* **2015**, *57*, 1248–1275. [[CrossRef](#)] [[PubMed](#)]
6. Meiring, G.A.; Myburgh, H.C. A Review of Intelligent Driving Style Analysis Systems and Related Artificial Intelligence Algorithms. *Sensors* **2015**, *15*, 30653–30682. [[CrossRef](#)] [[PubMed](#)]
7. Ericsson, E. Independent driving pattern factors and their influence on fuel-use and exhaust emission factors. *Transp. Res. Part D Transp. Environ.* **2001**, *6*, 325–345. [[CrossRef](#)]
8. Wijnands, J.S.; Thompson, J.; Aschwanden, G.D.P.A.; Stevenson, M. Identifying behavioural change among drivers using Long Short-Term Memory recurrent neural networks. *Transp. Res. Part F Traffic Psychol. Behav.* **2018**, *53*, 34–49. [[CrossRef](#)]
9. Ohn-Bar, E.; Trivedi, M.M. Looking at Humans in the Age of Self-Driving and Highly Automated Vehicles. *IEEE Trans. Intell. Veh.* **2016**, *1*, 90–104. [[CrossRef](#)]
10. Kuderer, M.; Gulati, S.; Burgard, W. Learning driving styles for autonomous vehicles from demonstration. In Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; pp. 2641–2646.
11. Warren, J.; Lipkowitz, J.; Sokolov, V. Clusters of Driving Behaviour from Observational Smartphone Data. *arXiv* **2017**.
12. Lin, N.; Zong, C.; Tomizuka, M.; Song, P.; Zhang, Z.; Li, G. An overview on study of identification of driver behaviour characteristics for automotive control. *Math. Probl. Eng.* **2014**, *2014*, 569109. [[CrossRef](#)]
13. Stenneth, L.; Thompson, K.; Stone, W.; Alowibdi, J.S. Automated transportation transfer detection using GPS enabled smartphones. In Proceedings of the 15th International IEEE Conference on Intelligent Transportation Systems, Anchorage, AK, USA, 16–19 September 2012.
14. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
15. Fitch, G.M.; Hanowski, R.J. Using naturalistic driving research to design, test and evaluate driver assistance systems. In *Handbook of Intelligent Vehicles*; Springer: London, UK, 2012; pp. 559–580.
16. Kusano, K.D.; Montgomery, J.; Gabler, H.C. Methodology for identifying car following events from naturalistic data. In Proceedings of the IEEE Intelligence Vehicles Symposium, Dearborn, MI, USA, 8–11 June 2014; pp. 281–285.
17. Castignani, G.; Derrmann, T.; Frank, R.; Engel, T. Driver Behaviour Profiling Using Smartphones: A Low-Cost Platform for Driver Monitoring. *IEEE Intell. Transp. Syst. Mag.* **2015**, *7*, 91–102. [[CrossRef](#)]
18. Zheng, X.; Chen, W.; Wang, P.; Shen, D.; Chen, S.; Wang, X.; Zhang, Q.; Yang, L. Big data for social transportation. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 620–630. [[CrossRef](#)]



19. Jariyasunant, J.; Carrel, A.; Ekambaram, V.; Gaker, D.J.; Kote, T.; Sengupta, R.; Walker, J.L. The Quantified Traveler: Using personal travel data to promote sustainable transport behavior. In Proceedings of the Transportation Research Board (TRB) 91st Annual Conference, Washington, DC, USA, 22–26 January 2012.
20. Ferrer, S.; Ruiz, T. Travel behaviour characterization using raw accelerometer data collected from smartphones. *Procedia Soc. Behav. Sci.* **2014**, *160*, 140–149. [[CrossRef](#)]
21. Engelbrecht, J.; Booysen, M.J.; van Rooyen, G.J.; Bruwer, F.J. Survey of Smartphone-Based Sensing in Vehicles for Intelligent Transportation System Applications. *IET Intell. Transp. Syst.* **2015**, *9*, 924–935. [[CrossRef](#)]
22. Nie, Y. How can the taxi industry survive the tide of ride sourcing? Evidence from Shenzhen, China. *Transp. Res. Part C Emerg. Technol.* **2017**, *79*, 242–256. [[CrossRef](#)]
23. Han, H.; Yu, J.; Zhu, H.; Chen, Y.; Yang, J.; Zhu, Y.; Xue, G.; Li, M. Sen-Speed: Sensing Driving Conditions to Estimate Vehicle Speed in Urban Environments. In Proceedings of the IEEE Transactions on Mobile Computing, Toronto, ON, Canada, 1 January 2016; Volume 15, pp. 202–216.
24. Carmona, J.; García, F.; Martín, D.; de la Escalera, A.; Armingol, J.M. Data Fusion for Driver Behaviour Analysis. *Sensors* **2015**, *15*, 25968–25991. [[CrossRef](#)] [[PubMed](#)]
25. Bohte, W.; Maat, K. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transp. Res. Part C Emerg. Technol.* **2009**, *17*, 285–297. [[CrossRef](#)]
26. Chen, C.; Gong, H.; Lawson, C.; Bialostozky, E. Evaluating the Feasibility of a Passive Travel Survey Collection in a Complex Urban Environment: Lessons learned from the New York City Case Study. *Transp. Res. Part A Policy Pract.* **2010**, *44*, 830–840. [[CrossRef](#)]
27. Karlaftis, M.G.; Vlahogianni, E.I. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transp. Res. Part C Emerg. Technol.* **2011**, *19*, 387–399. [[CrossRef](#)]
28. Breiman, L. Statistical modeling: The two cultures. *Stat. Sci.* **2001**, *16*, 199–231. [[CrossRef](#)]
29. Liu, H.; Taniguchi, T.; Tanaka, Y.; Takashi, B. Essential feature extraction of driving behaviour using a deep learning method. In Proceedings of the 2015 IEEE Intelligent Vehicles Symposium (IV), Seoul, Korea, 28 June–1 July 2015.
30. Zhao, C.; Zheng, C.; Zhao, M. Classifying driving mental fatigue based on multivariate autoregressive models and kernel learning algorithms. In Proceedings of the 3rd International Conference on Biomedical Engineering and Informatics, Yantai, China, 16–18 October 2010; Volume 6, pp. 2330–2334.
31. Engelbrecht, A.P. *Computational Intelligence*, 2nd ed.; An Introduction; Wiley: Hoboken, NJ, USA, 2007.
32. Schussler, N.; Axhausen, K.W. Processing raw data from global positioning systems without additional information. *Transp. Res. Rec. J. Transp. Res. Board* **2009**, *2105*, 28–36. [[CrossRef](#)]
33. Mozolin, M.; Thill, J.-C.; Lynn Usery, E. Trip distribution forecasting with multiplayer perceptron neural networks: A critical evaluation. *Transp. Res. Part B Methodol.* **2000**, *34*, 53–73. [[CrossRef](#)]
34. Tango, F.; Botta, M. Real-time detection system of driver distraction using machine learning. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 894–905. [[CrossRef](#)]
35. Wu, Y.; Tan, H. Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework. *arXiv* **2016**.
36. Yu, H.; Wu, Z.; Wang, S.; Wang, Y.; Ma, X. Spatiotemporal Recurrent Convolutional Networks for Traffic Prediction in Transportation Networks. *Sensors* **2017**, *17*, 1501. [[CrossRef](#)] [[PubMed](#)]
37. Van Ly, M.; Martin, S.; Trivedi, M. Driver classification and driving style recognition using inertial sensors. In Proceedings of the 2013 IEEE Intelligent Vehicles Symposium (IV), Gold Coast, QLD, Australia, 23–26 June 2013; pp. 1040–1045.
38. Dong, W.; Yuan, T.; Yang, K.; Li, C.; Zhang, S. Autoencoder Regularized Network for Driving Style Representation Learning. *arXiv* **2017**.
39. Kohonen, T. *Self-Organizing Maps*, 2nd ed.; Springer Series in Information Sciences; Springer: Berlin/Heidelberg, Germany, 1997.
40. Dorr, D.; Grabengieser, D.; Gauterin, F. Online driving style recognition using fuzzy logic. In Proceedings of the IEEE 17th International Conference on Intelligent Transportation Systems (ITSC), Qingdao, China, 8–11 October 2014; pp. 1021–1026.
41. Liu, L.; Andris, C.; Ratti, C. Uncovering cabdrivers' behaviour patterns from their digital traces. *Comput. Environ. Urban Syst.* **2010**, *34*, 541–548. [[CrossRef](#)]

42. Schwenk, H.; Milgram, M. Transformation invariant autoassociation with application to handwritten character recognition. In *Advances in Neural Information Processing Systems*; Tesauro, G., Touretzky, D., Leen, T., Eds.; MIT Press: Cambridge, MA, USA, 1995; Volume 7, pp. 991–998.
43. Sparse Autoencoder. Available online: <http://ufldl.stanford.edu/tutorial/unsupervised/Autoencoders/> (accessed on 20 December 2017).
44. Bengio, Y.; Le Cun, Y. Scaling learning algorithms towards AI. In *Large Scale Kernel Machines*; Bottou, L., Chapelle, O., DeCoste, D., Weston, J., Eds.; MIT Press: Cambridge, MA, USA, 2007.
45. Japkowicz, N.; Hanson, S.J.; Gluck, M.A. Nonlinear autoassociation is not equivalent to PCA. *Neural Comput.* **2000**, *12*, 531–545. [[CrossRef](#)] [[PubMed](#)]
46. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [[CrossRef](#)] [[PubMed](#)]
47. Lampinen, J.; Oja, E. Clustering properties of hierarchical self-organizing maps. *J. Math. Imaging Vis.* **1992**, *2*, 261–272. [[CrossRef](#)]
48. Vijaya Kumar, B.P.; Gazala, H.T. Self-organized mapping-based map-reduce technique in big data analytics: A neural network approach. *J. Eng. Appl. Sci.* **2016**, *11*, 18.
49. Soole, D.W.; Watson, B.C.; Fleiter, J.J. Effects of average speed enforcement on speed compliance and crashes: A review of the literature. *Accid. Anal. Prev.* **2013**, *54*, 46–56. [[CrossRef](#)] [[PubMed](#)]
50. Xu, Z.; Sun, J. Model-driven deep-learning. *Nat. Sci. Rev.* **2018**, *5*, 22–24. [[CrossRef](#)]
51. Yeo, N.C.; Lee, K.H.; Venkatesh, Y.V.; Ong, S.H. Colour image segmentation using the self-organizing map and adaptive resonance theory. *Image Vis. Comput.* **2005**, *23*, 1060–1079. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).