Drawing-Based Automatic Dementia Screening Using Gaussian Process Markov Chains

Max WY Lam[1], Xunying Liu[2], Helen ML Meng[1,2], Kelvin KF Tsoi[1,3]

[1]Stanley Ho Big Data Decision Analytics Research Centre
[2]Department of Systems Engineering and Engineering Management
[3]Jockey Club School of Public Health and Primary Care
The Chinese University of Hong Kong
kelvintsoi@cuhk.edu.hk

Abstract

Screening tests play an important role for early detection of dementia. Among those widely used screening tests, drawing tests have gained much attention in clinical psychology. Traditional evaluation of drawing tests totally relies on the appearance of drawn picture, but does not consider any time-dependent behaviour. We demonstrated that the processing speed and direction can reflect the decline of cognitive function, and thus may be useful for disease screening. We proposed a model of Gaussian process Markov chains (GPMC) to study the complex associations within the drawing data. Specifically, we modeled the process of drawing in a state-space form, where a drawing state is composed of drawing direction and velocity with consideration of the processing time. For temporal modeling, our scope focused more on discrete-time Markov chains on continuous state space. Because of the short processing time of picture drawing, we applied higher-order of Markov chains to model long-term temporal correlation across drawing states. Gaussian process regression was used for universal function approximation to flexibly infer the state transition function. With Gaussian process prior to the distribution of function space, we could encode high-level function properties such as noisiness, smoothness and periodicity. We also derived an efficient training mechanism for complex Gaussian process regression on bivariate Markov chains. With GPMC, we present an optimal decision rule based on Bayesian decision theory. We applied our proposed method to a drawing test for dementia screening, i.e. interlocking pentagon-drawing test. We tested our models with 256 *subjects who are aged from 65 to 95. Finally, comparing to the traditional methods, our models showed remarkable improvement in drawing test for dementia screening.*

1. Introduction

Drawing is a non-linguistic form of human expression of ideas. Digital drawing has been frequently discussed as a convenient interface for human-computer interaction. Thus, much effort in the research community has been devoted to recognize picture drawing [1-10], trying to maximize the recognition accuracy with state-of-the-art machine learning techniques. From another point of view, we can also take drawing process as a reflection of human cognition functions. In fact, drawing tests for disease screening are very common in the field of clinical psychology, where a number of validated drawing tests [11-16] have been applied in healthcare settings. Most of the tests are still conducted in paper-and-pencil form and relied on human decision by healthcare professionals, which usually involves subjective judgment. A decision to distinguish a straight line from a curve is a typical example.

Computerized evaluation provides an objective way to define a drawing picture being “good” or “bad”. Traditional methods evaluate a drawing picture with human subjective decision. Also, the motion of drawing is a complex factor that can be captured for further analysis. Many clinical findings suggested, the motion of drawing, such as tremor, is closely related to the symptoms of different types of dementia. While nowadays digital devices can already capture the drawing behaviour in human-unreachable details, we proposed to consider the motion of drawing data for dementia screening, and to recognize imperceptible drawing patterns that are crucial to distinguish dementia subjects from the general population. The challenge comes along the explicit consideration of drawing motion and can be a spatiotemporal model with an increased complexity of time.

URI: http://hdl.handle.net/10125/50241
ISBN: 978-0-9981331-1-9

In this study, we formulated the drawing motion as a discrete-time state-space model, in which each state is composed of two random variables representing drawing direction [1, 3-5, 9-10] and drawing velocity [1, 9-10]. Indeed, using state-space model is not a new concept in sketch recognition, but Markov models [2-6] and Bayesian networks [7] have been applied. The definition of training data also facilitates big data analytics. The collection of real-time drawing data is usually difficult to be analyzed by the traditional clinical data processing methods among healthcare professionals. We used one training sample for each drawing state, but not for each subject. Therefore, our models had more training samples to reduce the chance of over-fitting of the data.

The process of drawing was presented in a state-space form. In our formulation, each state in a drawing process is bivariate, encoding direction and velocity information within a short period of time. For temporal modeling, our scope focused on discrete-time Markov chains on continuous state space. Instead of the most commonly used first-order Markov chains, we proposed to use higher-order Markov chains because, for states defined in a short period of time, transitions across states tended to have higher-order temporal dependence.

To obtain a conditional likelihood in higher-order Markov chains on continuous state space, a general state transition function is necessary. A common way to learn the unknown state transition function is to define a parametric form, such as linear functions [17], and radial basis functions [20]. We did not restrict the state transition function to a class of mathematical functions parameterized by a finite set of parameters.

Instead, we placed a Gaussian process prior over an infinite-dimensional space of state transition function. This prior can encode function properties such as noisiness, smoothness and periodicity [25]. In fact, with an appropriate choice of the kernel function, the Gaussian process prior puts probability mass over all continuous functions [21]. The inference of the predictive distribution over the function space is renowned as Gaussian process regression model (GPR) [22]. However, exact inference of GPR is non-scalable. For practical concerns, we derived an efficient complex GPR for bivariate Markov chains.

In this paper, GPR is first to be proposed as a general solution to find state transition function of conditional higher-order discrete-time Markov chains on continuous state space. We put this generic modeling approach as Gaussian process Markov chains (GPMC). As a generative model, GPMC can be used to infer the probability of having dementia given drawing behavioural data. Since GPMC is derived from Bayesian framework, we came up with an optimal decision rule based on Bayesian decision theory. To minimize the chance of false negative results, we put more allowance for the false positive results. We applied our proposed method to a drawing screening test, namely, interlocking pentagon drawing test. Finally, we tested our models with 256 subjects aged from 65 to 95, and compared the result to traditional screening tests. Our method reported a remarkable improvement over the previous evaluation schemes for interlocking pentagon drawing test.

In the sections after this introduction, we begin to introduce our modeling approach for digital drawing in section 2. We formally define the task of drawing-based for dementia screening, and then derive GPMC to deal with the task in section 3. We compare the performance with experimental evidence in section 4. Finally, we conclude our work in section 5.

2. Modeling for Digital Drawing

This section is structured as follows: in section 2.1, we define the structure of raw drawing data collected from digital devices; in section 2.2, we formulate a state-space representation for digital drawing with clinical interpretation; in section 2.3, we define statistical properties by linking our state-space representation to Markov Chains.

2.1. Raw Data of Digital Drawing

For any subject denoted by $\boldsymbol{S}$, the drawing process $D(\mathbf{S})$ is composed of a sequence of strokes $s_1, \dots, s_L$, preserving the drawing order, while a stroke $s_m$contains a sequence of points $p_{i,1}, \dots, p_{i,M_i}$. At each point $p_{i,j}$, we capture three quantities:

$$\left(x_{i,j}, y_{i,j}, t_{i,j}\right),$$

where $x_{i,j}, y_{i,j}$ corresponds to the coordinate of the drawing point, and $t_{i,j} \in \mathbb{R}^+$ is the amount of seconds spent on the point $p_{i,j}$ before moving to the next point $p_{i,j+1}$. Combining above definitions, we simply denote a drawing process by

$$D(\mathbf{S}) = \left[\left(x_r, y_r, t_r, e_r\right)\right]_{r=1}^{N},$$

where e_r is a binary indicator denoting whether the rth point is at the end of the stroke, and $N = \sum_{i=1}^{L} M_i$ is the total number of points captured in the drawing process. It is notable that for different devices the density of illuminated points could also be varied. As a result, it is better to convert pixels to centimeters with reference to pixels per inch (ppi).

2.2. State-Space Representation

Having defined the raw data structure for digital drawing, we proceeded by defining higher-level features that are more expressible to human, and are useful for inference. Two quantities from the raw data were extracted, including: i) drawing direction θ_r [1, 3-5, 9-10] and ii) drawing velocity v_r[1, 9-10]. Formally, for $r \in \{1, \dots, N\}$, if $e_r = 0$,

$$\theta_r = \tan^{-1}\left(\frac{y_{r+1} - y_r}{x_{r+1} - x_r}\right),$$

$$v_r = \frac{\sqrt{(y_{r+1} - y_r)^2 + (x_{r+1} - x_r)^2}}{t_r},$$

otherwise, if $e_r = 1$, the point is at the end of a stroke, therefore θ_r and v_r are not defined.

In fact, drawing direction and velocity constitute a natural description of drawing motion. Note that we did not consider the locations of image on the drawing panel which are highly varied across different subjects. For example, a subject can put his/her drawing picture on a corner of the drawing panel. Our models are more robust if we only considered two quantities in a state for direction and velocity. In this study, we investigated the nature of these two features, and discuss the connection to clinical findings.

2.2.1. Feature of Drawing Direction

In an ideal case, direction of a single straight line can be directly associated with the drawing time, but the movement of a stroke can be curve-shaped that we need to measure the direction changes in angles. Angles would remain unchanged for drawing a perfect straight line. Therefore, angular feature is capable of reflecting information such as line straightness, corner sharpness, and tremors. We can also capture information that associates with neuropsychological findings, such as disability to draw horizontal line [40], chorea (involuntary movement), and akinesia (difficulty in maintaining voluntary movement), which had been shown to be particularly sensitive to Huntington's disease (HD) [41] and Parkinson's disease (PD) [42]. In particular, HD had been reported to cause a degeneration of striatum [43], which would result in bradykinesia (difficulty in maintaining movement) [44]. Similarly, PD is associated with a loss of dopaminergic cells in the substantia nigra that associated with the striatum, which lead to disturbances of motor control [45]. Besides HD and PD, clinical studies had also found a degradation of motor program in other common types of dementia like Alzheimer's disease (AD) [46] and dementia with Lewy Bodies (DLB) [47], while its impact on tremors or other involuntary movements can be analyzed through angular feature [48].

2.2.2. Feature of Drawing Velocity

Velocity is a quantity due to its unique physical meaning. In fact, HD and PD patients, who suffered from bradykinesia, revealed difficult to produce rapid voluntary motor activity [49]. As an example, reach-to-grasp tasks [50] showed that motions of PD patients are 30% slower than the control group. Furthermore, disorders of the visual ability as well as constructional disorganization could induce a longer drawing time and a lower velocity. This kind of disorders can be found commonly in AD [51-52] and DLB [53-54].

To enable comparison across different drawing forms and processes, we used a discrete-time state-space representation, where a drawing process is divided into n time blocks of equal width, yielding n drawing states. Each drawing state, denoted by $\mathbf{D}_i$, encodes drawing direction and drawing velocity within the time block. Formally, we define $\mathbf{D}_i = [\boldsymbol{\theta}_i, \boldsymbol{v}_i]$, where $\boldsymbol{\theta}_i$ and $\boldsymbol{v}_i$ are random variables that describe how a subject draws a stroke in terms of direction and velocity within the ith time block. For the ease of explanation, in the following parts of this paper, the shorthand $\mathbf{D}_{a:b}$ was used to denote the sequence of state $\mathbf{D}_a, \dots, \mathbf{D}_b$.

2.3. Markov Chains for Drawing Processes

Traditional screening on a drawing test usually concerns about the shape of final drawn picture, i.e., spatial information, but we conjectured that imperceptible drawing behaviors can be captured with the time, i.e., spatial-temporal information. In fact, spatial-temporal modeling is not a new concept in sketch recognition communities, where hidden Markov models [2-6] and Bayesian networks [7] have been used. It is advantageous to refer to the successful modeling approaches in sketch recognition. Yet, these methods cannot be directly applied to drawing-based screening tests. In addition to recognizing the final drawn picture, we are trying to capture the intermediate cognitive reflection during the drawing process such as delays, redrawing, or peculiar movements.

In particular, we put emphasis on temporal information. Due to time dependency in the drawing task, it is intuitive to claim that a state $\mathbf{D}_i$ is dependent on its preceding states $\mathbf{D}_{1:i-1}$, and is

completely unaffected by the future states $\mathbf{D}_{i+1:n}$. We formally express this reasoning with statistical basis:

$$p(\mathbf{D}_{1:i}) = \prod_{j=1}^{i} p(\mathbf{D}_j|\mathbf{D}_{1:j-1}).$$

Although the future states are neglected, there are still a long chain of conditionals that involves messy calculations. A common choice of available solutions is to assume Markov property [2-6], which asserts that the next state only depends on nearby states, and is conditionally independent of the previous states. For nowadays applications of Markov models, first-order Markov property [18], which mentions that the probability distribution of future state is dependent only upon the present state, is the most frequently used, but is not likely satisfied in our case. Therefore, we relaxed the assumption to allow the future state depending on the past m states. This is also known as a higher-order Markov chain, which is very useful as a mathematical tool [19]. Probabilistically, we now approximate the joint probability by

$$p(\mathbf{D}_{1:i}) \approx \prod_{j=1}^{i} p(\mathbf{D}_j|\mathbf{D}_{j-m:j-1}).$$

3. Drawing for Dementia Screening

In this section, we present a model to infer the probability of having dementia given drawing data for dementia screening. In section 3.1, we define the problem probabilistically. In section 3.2, we define the prior. In section 3.3, we derive Gaussian process Markov chains (GPMC) to find the likelihood. To cope with bivariate Markov chains, in section 3.4 we present a complex-valued GPR. In section 3.5, we show an optimal decision rule based on GPMC.

3.1. Problem Definition

Our goal is to decide whether a subject with certain degree of cognitive impairment (CI) can be identified with reference to the process of drawing. Instead of classifying the subjects into dichotomous outcomes of cognitive healthy or unhealthy, the fuzzy logic approach is more appropriate to describe the chance of have CI as a continuous value from 0 to 1. Naturally, we would make use of probability as a well-established tool to represent such a value.

With our aforementioned state-space concepts, our task becomes to find the likelihood of a sequence of drawing states being generated by a subject with CI. Formally, we use $\mathbf{C} = \mathrm{c} \in \{1, 0\}$ to denote whether the subject have CI or not. In addition to the drawing performance, it is sensible to consider other factors that may affect the chance of having CI. In this paper, we additionally considered age $\mathbf{A}$, gender $\mathbf{G}$, and education level $\mathbf{E}$ [32-35]. Generally, we used $\mathbf{B}$ to denote the subject's background information that can be collected together with the drawing data.

From a probabilistic perspective, our target can be written as $p(\mathbf{C}|\mathbf{D}_{1:n}, \mathbf{B})$, so that we can theoretically give a screening result by comparing $p(\mathbf{C} = 1|\mathbf{D}_{1:n} = \mathbf{d}_{1:n}, \mathbf{B} = \mathbf{b})$ to $p(\mathbf{C} = 0|\mathbf{D}_{1:n} = \mathbf{d}_{1:n}, \mathbf{B} = \mathbf{b})$, where $\mathbf{d}_{1:n}$ is the vector corresponding to observe drawing states, and $\mathbf{b}$ is the vector corresponding to subject's background. In our case, $\mathbf{b} = [a, g, e]$, providing that a is the age of the subject, g is the gender of the subject, e is the education level of the subject. From Bayes' rule, we know that posterior is proportional to the prior times the likelihood:

$$p(\mathbf{C}|\mathbf{D}_{1:n}, \mathbf{B}) \propto p(\mathbf{B}, \mathbf{C})p(\mathbf{D}_{1:n}|\mathbf{B}, \mathbf{C}).$$

Since the evidence $p(\mathbf{D}_{1:n}, \mathbf{B})$ is not affected by $\mathbf{C}$, the problem is reduced to finding the prior $p(\mathbf{B}, \mathbf{C})$ and the likelihood $p(\mathbf{D}_{1:n}|\mathbf{B}, \mathbf{C})$.

3.2. Prior Distribution: Maximum Likelihood Estimation

To reasonably set up a prior distribution $p(\mathbf{B}, \mathbf{C})$, we take advantage of the conditional dependence:

$$p(\mathbf{B}, \mathbf{C}) = p(\mathbf{B}|\mathbf{C})p(\mathbf{C}),$$

where the conditional probability $p(\mathbf{B}|\mathbf{C})$, in our case, equals to $p(\mathbf{A}|\mathbf{C})p(\mathbf{G}|\mathbf{C})p(\mathbf{E}|\mathbf{C})$. Presumably, the conditional variables follow below distributions:

$$\mathbf{A}|\mathbf{C} \sim \mathcal{N}(\mu_{\mathrm{age}|\mathbf{C}}, \sigma^2_{\mathrm{age}|\mathbf{C}}),$$
$$\mathbf{G}|\mathbf{C} \sim \mathrm{Ber}(p_{\mathrm{male}|\mathbf{C}}),$$
$$\mathbf{E}|\mathbf{C} \sim \mathrm{Cat}(p_{\mathrm{ued}|\mathbf{C}}, p_{\mathrm{pri}|\mathbf{C}}, p_{\mathrm{sec}|\mathbf{C}}, p_{\mathrm{uni}|\mathbf{C}}),$$

where $\mu_{\mathrm{age}|\mathbf{C}}$ and $\sigma^2_{\mathrm{age}|\mathbf{C}}$ are the mean and the variance of a subject's age respectively in a Gaussian distribution, $p_{\mathrm{male}|\mathbf{C}}$ is the probability of being a male subject in a Bernoulli distribution, $p_{\mathrm{ued}|\mathbf{C}}$, $p_{\mathrm{pri}|\mathbf{C}}$, $p_{\mathrm{sec}|\mathbf{C}}$, and $p_{\mathrm{uni}|\mathbf{C}}$ are the probabilities of being an uneducated subject, being a subject graduated from primary school, being a subject graduated from secondary school, and being a subject graduated from university school respectively in a categorical distribution. Note that we have two sets of parameters corresponding two possible outcomes of $\mathbf{C}$. To calculate the parameters, we used maximum likelihood estimation (MLE) in the training data.

On the other hand, for $p(\mathbf{C})$, we similarly assume $\mathbf{C} \sim \mathrm{Ber}(p_{\mathrm{ci}})$, but it is usually problematic to directly estimate p_{ci} from the training data as we often set selection criteria to control the number of dementia patient while collecting the training samples. Therefore, we employed random sampling of a larger pool to estimate p_{ci}. In a realistic setup,

the prior belief addressed by p_{ci} should be varied over counties. In our work, we referred to the statistics reported by a clinical study [58], suggesting $p_{\text{ci}} = 0.145$.

3.3. Likelihood Distribution: Gaussian Process Markov Chains

In the last section, we had shown that $p(\mathbf{B}, \mathbf{C})$ can be estimated from statistical inference. This part is to infer $p(\mathbf{D}_{1:n}|\mathbf{B}, \mathbf{C})$, which accounts for the likelihood of drawing being produced by a subject with or without CI from different background. The solution is closely related to our state-space modeling approach which was introduced in section 2. Specifically, we considered the higher-order Markov chains in conditional form, i.e., specifying modeling conditions such that $p(\mathbf{D}_{1:n})$becomes $p(\mathbf{D}_{1:n}|\mathbf{B}, \mathbf{C})$. A typical practice is to independently train several Markov chains with the corresponding subsets of data. However, this approach usually requires a large training dataset to prevent over-fitting of the model [18]. The problem of over-fitting may become severe when a Markov chain is in a higher order, where the number of parameters increases exponentially with the order of Markov chain. An order-m Markov chain taking values in a finite set of size k has $k^m(k-1)$ independent transition probabilities [19]. In our case, $\mathbf{D}_i$ takes values in an uncountable set. Apparently, traditional formulation of finite-state Markov chains is infeasible.

To overcome the hurdles for conditional Markov chains on continuous state space, we defined a general state transition function $f(\cdot)$, which presumably can capture the transition dependencies with some tolerable noise, i.e.,

$$\boldsymbol{y}_j = f(\mathbf{x}_j) + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2)$$

where

$$\mathbf{x}_j = [\mathbf{d}_{j-m:j-1}, \mathbf{b}, c]$$

is a D-dimensional input vector that represents previous m states, subject's background and CI condition, and

$$\boldsymbol{z}_j = z(\mathbf{D}_j) = \boldsymbol{v}_j + i\boldsymbol{\theta}_j$$

is a complex-valued random variable that represent a bivariate drawing state $\mathbf{D}_j$. Here, $z(\cdot)$ maps a two-dimensional random vector to a complex-valued random variable. A parametric approach to find $f(\cdot)$ consists in specifying a class of mathematical functions parameterized by a finite set of parameters. However, we used a less restrictive approach to infer $f(\cdot)$ by directly specifying a prior over an infinite-dimensional space of functions, i.e., a Gaussian process prior. Specifically, we define

$$f(\cdot) \sim \mathcal{GP}(0, k(\cdot,\cdot)),$$

which is a zero-mean Gaussian process that is fully specified by a kernel function $k(\cdot,\cdot)$. There are many analytical properties of $k(\cdot,\cdot)$. By choosing a higher-level parametric function $k(\cdot,\cdot)$ we encoded properties such as noisiness, smoothness and periodicity [25]. Neal [26] has proved that Bayesian neural networks with infinitely many hidden units converged to a Gaussian process with particular kernel function. Having defined a prior over $f(\cdot)$, we need to update the distribution of $f(\cdot)$ with training data.

Given the training set $\mathfrak{D} = \{\boldsymbol{z}_i, \mathbf{x}_i\}_{i=1}^{n(S-1)}$ that contains the drawing data of $S-1$ subjects, we denoted the training targets by $\boldsymbol{z} = [\boldsymbol{z}_i]_{i=1}^{n(S-1)}$, and the training inputs by $\boldsymbol{x} = [\mathbf{x}_i]_{i=1}^{n(S-1)}$. A finite collection of function variables $\boldsymbol{\mathcal{F}} = [f(\mathbf{x}_i)]_{i=1}^{n(S-1)}$ was created corresponding to the training inputs. By the definition of Gaussian process, $\boldsymbol{\mathcal{F}}$ follows multivariate Gaussian: $p(\boldsymbol{\mathcal{F}}) = \mathcal{N}(\mathbf{0}, \mathbf{K})$,
where $\mathbf{K}$ is computed from the kernel function:

$$\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j).$$

From our definition of state transition function, we have

$$p(\boldsymbol{z}|\boldsymbol{\mathcal{F}}) = \mathcal{N}(\boldsymbol{\mathcal{F}}, \sigma^2\mathbf{I}).$$

Using Bayes' rule,

$$p(\boldsymbol{\mathcal{F}}|\boldsymbol{z}) = \frac{p(\boldsymbol{z}|\boldsymbol{\mathcal{F}})p(\boldsymbol{\mathcal{F}})}{p(\boldsymbol{z})},$$

where the denominator is a constant, we can update the prior of $\boldsymbol{\mathcal{F}}$ and get the posterior:

$$p(\boldsymbol{\mathcal{F}}|\boldsymbol{z}) = \mathcal{N}\begin{pmatrix}\mathbf{K}(\mathbf{K}+\sigma^2\mathbf{I})^{-1}\boldsymbol{z}, \\ \mathbf{K}-\mathbf{K}(\mathbf{K}+\sigma^2\mathbf{I})^{-1}\mathbf{K}\end{pmatrix}.$$

Then, for an unseen input $\mathbf{x}^*$, we want to find the predictive distribution derived from the posterior:

$$p(f(\mathbf{x}^*)|\boldsymbol{z}) = \int p(f(\mathbf{x}^*)|\boldsymbol{\mathcal{F}})p(\boldsymbol{\mathcal{F}}|\boldsymbol{z})d\boldsymbol{\mathcal{F}}$$

where $p(f(\mathbf{x}^*)|\boldsymbol{\mathcal{F}})$ results from the Gaussian process prior linking all possible values of $\boldsymbol{\mathcal{F}}$ and $f(\mathbf{x}^*)$ with a joint Gaussian distribution. Finally, we can obtain

$$p(f(\mathbf{x}^*)|\boldsymbol{z}) = \mathcal{N}\begin{pmatrix}\mathbf{k}^*(\mathbf{K}+\sigma^2\mathbf{I})^{-1}\boldsymbol{z}, \\ k(\mathbf{x}^*,\mathbf{x}^*) - \mathbf{k}^*(\mathbf{K}+\sigma^2\mathbf{I})^{-1}\mathbf{k}^{*\mathrm{T}}\end{pmatrix},$$

where $\mathbf{k}^* = [k(\mathbf{x}^*, \mathbf{x}_i)]_{i=1}^{n(S-1)}$ is the kernel vector.

This exact inference procedure is popularized as GPR [22]. Eventually, by fitting the predictive distribution given by GPR into higher-order Markov chain, we obtained the joint likelihood of drawing being produced by a dementia subject:

$$p(\mathbf{D}_{1:n}|\mathbf{B}, \mathbf{C}) \approx \prod_{j=1}^{n} p\left(z(\mathbf{D}_j)|f(\mathbf{x}_j)\right) p\left(f(\mathbf{x}_j)|\boldsymbol{z}\right).$$

This approach is defined as GPMC, which generally works for conditional higher-order discrete-time Markov chains on continuous state space. For a

generic setting where states are not bivariate, we can set $\boldsymbol{z}_j = \mathbf{D}_j$.

3.4. Complex Gaussian Process Regression

In fact, two practical problems in GPMC were still unsolved for the task of drawing to distinguish dementia subjects: i) inversion of $\mathbf{K} + \sigma^2\mathbf{I}$ involves $\mathcal{O}(n^3S^3)$ complexity which is not scalable, and ii) complex-valued target is seldom discussed with GPR, which probably involves inversion of complex kernel matrix that is not Hermitian positive-definite. In this regard, we extended the standard GPR to complex-valued regression using the sparse spectrum approximation method [23]. Our extended approach shares the benefits of the sparse spectrum approximation method, where kernel function can be optimally found from the training data, and, more importantly, the training procedures are more efficient.

The idea is called sparse spectrum approximation of Gaussian process regression (SSGPR) [23], which in fact is a special case of randomized feature space for kernel machines [30]. The starting point of derivation is to assume a stationary kernel function:

$$k(\mathbf{x}_i, \mathbf{x}_j) = k(\mathbf{x}_i - \mathbf{x}_j).$$

Then, following Wiener-Khintchine theorem [28-29], the power spectral density and the autocorrelation function of a stationary random process together constitute a Fourier pair:

$$k(\mathbf{x}_i - \mathbf{x}_j) = \int_{\mathbb{R}^{2m}} S(\boldsymbol{\omega}) e^{2\pi i \boldsymbol{\omega}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j)} d\boldsymbol{\omega}.$$

From Bochner's theorem [30], which states that a stationary kernel function can be represented as Fourier transform of a positive finite Borel measure, we know that $S(\boldsymbol{\omega})$ is directly proportional to certain probability measure:

$$S(\boldsymbol{\omega}) \propto p(\boldsymbol{\omega}) \Leftrightarrow S(\boldsymbol{\omega}) = \eta^2 p(\boldsymbol{\omega}).$$

Combining above theorems, we get

$$k(\mathbf{x}_i - \mathbf{x}_j) = \sigma_k^2 \mathbb{E}_{p(\boldsymbol{\omega})}\left[e^{2\pi i \boldsymbol{\omega}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j)}\right].$$

The heart of SSGPR is to approximate the real part of this expectation with Monte-Carlo simulation [31]. The resultant becomes a sum of inner product:

$$\mathbb{E}_{p(\boldsymbol{\omega})}\left[\mathrm{Re}\left\{e^{2\pi i \boldsymbol{\omega}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j)}\right\}\right] \approx \frac{1}{K}\sum_{k=1}^{K} \cos\left(2\pi \hat{\boldsymbol{\omega}}_k{}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j)\right)$$

by which the kernel matrix can be equivalently expressed as a matrix-matrix multiplication:

$$\mathbf{K} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi},$$

given that

$$\boldsymbol{\Phi} = \frac{\eta}{\sqrt{K}}\begin{pmatrix} \sin(2\pi\hat{\boldsymbol{\omega}}_1{}^{\mathrm{T}}\mathbf{x}_1) & \cdots & \sin(2\pi\hat{\boldsymbol{\omega}}_1{}^{\mathrm{T}}\mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ \sin(2\pi\hat{\boldsymbol{\omega}}_K{}^{\mathrm{T}}\mathbf{x}_1) & \cdots & \sin(2\pi\hat{\boldsymbol{\omega}}_K{}^{\mathrm{T}}\mathbf{x}_n) \\ \cos(2\pi\hat{\boldsymbol{\omega}}_1{}^{\mathrm{T}}\mathbf{x}_1) & \cdots & \cos(2\pi\hat{\boldsymbol{\omega}}_1{}^{\mathrm{T}}\mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ \cos(2\pi\hat{\boldsymbol{\omega}}_K{}^{\mathrm{T}}\mathbf{x}_1) & \cdots & \cos(2\pi\hat{\boldsymbol{\omega}}_K{}^{\mathrm{T}}\mathbf{x}_n) \end{pmatrix},$$

is a $2K \times n$ feature matrix.

By Woodbury formula, $(\mathbf{K} + \sigma^2\mathbf{I})^{-1}$ in SSGPR becomes

$$(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} + \sigma^2\mathbf{I})^{-1} = \sigma^{-2}(\mathbf{I} - \boldsymbol{\Phi}^{\mathrm{T}}\mathbf{A}^{-1}\boldsymbol{\Phi}),$$

where $\mathbf{A} = \boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}} + \sigma^2\mathbf{I}$, is a $2\mathrm{K} \times 2\mathrm{K}$ matrix. Therefore, the time complexity of $(\mathbf{K} + \sigma^2\mathbf{I})^{-1}$ is reduced to $\mathcal{O}(Kn^2 + K^3)$, for a smaller $K \ll n$ [23].

Note that the initiative of SSGPR is to speed up the standard GPR training. Thus, the authors used only the real part of the complex exponential to resemble the originally-defined real-valued GPR. In fact, by this derivation, we can also come up with an analytical setup of complex GPR, which at the same time is as efficient as SSGPR. Our approach is to approximate the full expectation with Monte-Carlo simulation, i.e.,

$$\mathbb{E}_{p(\boldsymbol{\omega})}\left[e^{2\pi i \boldsymbol{\omega}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j)}\right] \approx \frac{1}{K}\sum_{k=1}^{K} e^{2\pi i \hat{\boldsymbol{\omega}}_k{}^{\mathrm{T}}\mathbf{x}_i}\left(e^{2\pi i \hat{\boldsymbol{\omega}}_k{}^{\mathrm{T}}\mathbf{x}_j}\right)^*,$$

where $(\,)^*$ denotes the complex conjugate. Similarly, we now have a $K \times n$ feature matrix

$$\boldsymbol{\Phi} = \frac{\eta}{\sqrt{K}}\begin{pmatrix} e^{2\pi i \hat{\boldsymbol{\omega}}_1{}^{\mathrm{T}}\mathbf{x}_1} & \cdots & e^{2\pi i \hat{\boldsymbol{\omega}}_1{}^{\mathrm{T}}\mathbf{x}_n} \\ \vdots & \ddots & \vdots \\ e^{2\pi i \hat{\boldsymbol{\omega}}_K{}^{\mathrm{T}}\mathbf{x}_1} & \cdots & e^{2\pi i \hat{\boldsymbol{\omega}}_K{}^{\mathrm{T}}\mathbf{x}_n} \end{pmatrix}.$$

The complex-valued kernel matrix becomes

$$\mathbf{K} = \boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi},$$

where $\boldsymbol{\Phi}^{\mathrm{H}}$ is the conjugate transpose of $\boldsymbol{\Phi}$. Also,

$$\mathbf{A} = \boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{H}} + \sigma^2\mathbf{I}.$$

With this formulation, the predictive distribution can simplified as

$$p(f(\mathbf{x}^*)|\boldsymbol{z}) = \mathcal{N}\begin{pmatrix} \boldsymbol{\phi}^{*\mathrm{H}}\mathbf{A}^{-1}\boldsymbol{\Phi}\boldsymbol{z}, \\ \sigma^2\boldsymbol{\phi}^{*\mathrm{H}}\mathbf{A}^{-1}\boldsymbol{\phi}^* \end{pmatrix},$$

where $\boldsymbol{\phi}^* = \frac{\eta}{\sqrt{K}}\left[e^{2\pi i \hat{\boldsymbol{\omega}}_k{}^{\mathrm{T}}\mathbf{x}^*}\right]_{k=1}^{K}$ is the feature mapping applied to the prediction inputs $\mathbf{x}^*$.

Based on the above derivation, we implemented a Python module for complex Gaussian process regression called GomPlex, [1] which was tested in IPython notebook on the cloud platform provided by IBM Data Science Experience. [2] The advantage of

[1] GomPlex: Complex Gaussian Process Regression - https://github.com/MaxInGaussian/GomPlex

[2] https://apsportal.ibm.com/analytics/notebooks/235da738-5c2d-4509-868f-7c1b4c3dccb9/view?access_token=cfaa2bfe988432d6137899369dd12f1cbc69b2d2d0b6278e318c81ede0429d3c

using this platform is mainly on the convenience of sharing results among interdisciplinary research team.

3.5. Decision Rules for Dementia Screening

Let w_+ be the decision of positive screening result, and w_- be the decision of negative screening result. Note that these two decisions may not be equally good or costly. We defined $\mathcal{C}(w, c)$ to be the cost of choosing $w \in \{w_+, w_-\}$ while the subject's status of dementia is $\mathbf{C} \in \{0, 1\}$. In this sense, $\mathcal{C}(w_+, 0)$ is the cost of false positive, while $\mathcal{C}(w_-, 1)$ is the cost of false negative. Based on Bayesian decision theory [36], by letting $\pi_{\mathrm{CI}} = p(\mathbf{C} = 1|\mathbf{D}_{1:n}, \mathbf{B})$ be the probability of having CI given the drawing data and subject's background, we obtained the expected risks for choosing the two decisions:

$$\mathcal{R}(w_+) = \mathcal{C}(w_+, 1)\pi_{\mathrm{CI}} + \mathcal{C}(w_+, 0)(1 - \pi_{\mathrm{CI}}),$$
$$\mathcal{R}(w_-) = \mathcal{C}(w_-, 1)\pi_{\mathrm{CI}} + \mathcal{C}(w_-, 0)(1 - \pi_{\mathrm{CI}}).$$

Essentially, for a screening test, false negative [37] is critical as early prevention is prohibited, and adverse events may be generated because of underestimation. In dementia screening, as there are currently no specific treatments to block the progression of cognitive decline in dementia [38]. Early detection allows the early plan of treatment or interventions [39]. In contrast, the cost of false positive equals to the drawbacks of preventive treatment, which is comparatively harmless. As a matter of fact, we set

$$\mathcal{C}(w_-, 1) = (1 + \delta) \cdot \mathcal{C}(w_+, 0),$$

where $\delta \in \mathbb{R}^+$ is a parameter pre-specified to the system that asserts $\mathcal{C}(w_-, 1) \geq \mathcal{C}(w_+, 0)$. Since we normally embrace correct predictions, vanishing cost

$$\mathcal{C}(w_+, 1) = \mathcal{C}(w_-, 0) = 0.$$

Above specifications lead to a simplified form of expected risks:

$$\mathcal{R}(w_+) = \mathcal{C}(w_+, 0)(1 - \pi_{\mathrm{CI}}),$$
$$\mathcal{R}(w_-) = \mathcal{C}(w_+, 0)(1 + \delta)\,\pi_{\mathrm{CI}}.$$

The updated optimal decision rule becomes

$$choose \begin{cases} w_+, & \text{if } \pi_{\mathrm{CI}} \geq \dfrac{1}{2 + \delta}, \\ w_-, & \text{if } \pi_{\mathrm{CI}} < \dfrac{1}{2 + \delta}. \end{cases}$$

4. Application: Pentagon Drawing Test

In this section, we show an application of our proposed modeling method to a drawing test – interlocking pentagon drawing test, which has been shown to be correlated with the measures of visuospatial abilities, memory and attention [14]. The details of data collection and evaluation results are shown as follow:

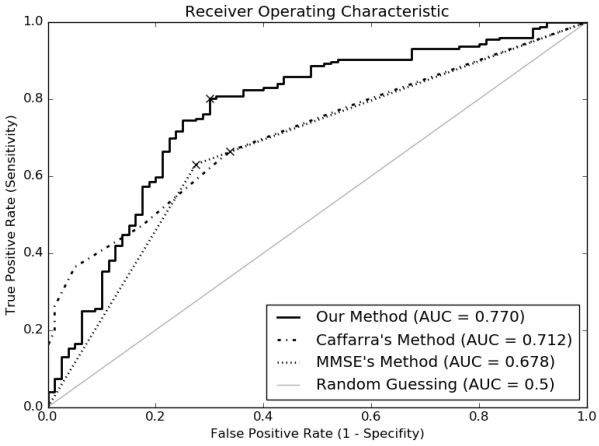

Figure 1. ROC Curve of different methods [15-16] for pentagon-copying test

4.1. Data Collection

In our experiment, 256 subjects were recruited to draw the interlocking pentagons on our digital platform. Participants drew two overlapping pentagons on a touchscreen of an Android tablet with reference to a sample figure. Among the 256 subjects, 44 subjects were recruited from the dementia clinics and diagnosed with moderate-to-severe stage of Alzheimer's Disease (AD), while 212 subjects aged 65 or above were recruited from the community without clinical symptoms of dementia. Prior to the test, all subjects had assessed by the Montreal Cognitive Assessment (MoCA) test [24], which is widely used for dementia screening. Therefore, for the 212 participants from the community, we used their MoCA score as the indicator of dementia and used a cut-off of 21 for CI, according to a local study in Hong Kong [55]. From this criterion, 132 out of 212 community participants were labeled as potential cases of dementia. Therefore, in this study, a total of 176 participants were classified as dementia subjects and 80 participants were healthy subjects.

4.2. Evaluation Results

The performance of our computational methods was compared to the traditional paper-and-pencil methods on the same group of subjects. The traditional scoring method was mainly based on the Mini-Mental State Examination (MMSE) [15], where the scoring guideline for frontier physicians is concluded into one statement: "the subject must draw two 5-sided figures intersected by a 4-sided figure". If this statement is violated, then the subject will receive 0 score, otherwise he/she will receive 1 score.

Table 1. Screening Performance of Different Dementia Screening Methods

	MMSE's Method	Caffarra's Method	Our Method
AUC	0.6778	0.7115	**0.7702**
Sensitivity	63.07%	66.48%	**80.11%**
Specificity	**72.5%**	66.25%	70.00%
Precision	83.46%	81.25%	**85.45%**
F1 Score	71.84%	73.13%	**82.70%**
Accuracy	66.02%	66.41%	**76.95%**

The binary scoring between 0 and 1 on the pentagon drawing test is not favorable to represent the visuospatial performance, thus a new qualitative scoring method for MMSE pentagon test has been proposed by Caffarra [16]. We referred it as another reference method. Caffarra's method included five factors of consideration: i) the numbers of angles, ii) distance/ intersection between the two figures, iii) closing/opening of the contour, iv) rotation of one or both pentagons, and v) closing-in.

Regarding our proposed method, we completed out-of-sample evaluation with leave-one-out cross-validation. That is, to calculate $\pi_{CI}$for a person, that person's drawing data as well as background information are excluded from GPMC totally.

Receiver Operating Characteristic (ROC) curve is a plot of true positive rate (sensitivity) versus false positive rate (1-specificity) that plays a prime role in evaluating diagnostic tests and finding the optimal cut-off in medical research [56]. Meanwhile, since our derived optimal decision is determined by the cut-off of the probability measure π_{CI}, i.e., the pre-specified parameter δ, it is well-suited to analyze the predictive performances across all possible thresholds through a ROC curve. The ROC curve comparing the performance of all methods is shown in Figure 1.

By maximizing sensitivity plus specificity, we also found optimal cut-offs for all methods, marked with crosses in the figure. The performances of all methods corresponding their optimal cut-offs are listed on Table 1. We calculated the most commonly used quantitative measures for classification problem including the area under the ROC curve (AUC) [57], sensitivity, specificity, precision, F1 score, and accuracy. Our computational method achieves the best among 4 measures out of 5 measures. All in all, with the optimal cut-off, our method has shown superiority over the two methods. Besides the predictive performance, the time-dependent behaviour is also a considerable factor for dementia screening. In the 256 participants, the average drawing time was 58.1 seconds to copy the overlapping pentagons, whereas traditional screening test, such as MoCA [24], takes at least 5 to 10 minutes for the whole evaluation. In this regard, our system is cost-effective for a population-based screening.

5. Conclusions

Digital drawing is potentially a valuable solution for dementia screening. It is demonstrated to be a simple and effective test. Although image recognition is common on digital drawing test, time-dependent behaviour is more directly related to the cognitive functions during the drawing. In this paper, we constituted three contributions. The first one is to show a novel state-space representation that can be used to quantify a drawing process. Robustness and connection to clinical studies have been found for our formulation. The second contribution, which is also the highlight of this paper, is to introduce Gaussian process Markov Chain (GPMC), by which we estimate the joint likelihood of conditional higher-order discrete-time Markov chains on continuous state space. The third contribution of this paper is to derive an efficient algorithm for complex Gaussian process regression. This allows us to generalize transition probabilities in Markov chain composed of bivariate states. Finally, we deduced an optimal decision rule from Bayesian decision theory.

6. References

[1] H. S. Yoon et al., "Hand gesture recognition using combined features of location, angle and velocity", Pattern Recognition, 34(7), 2001, pp. 1491-1501.

[2] S. Simhon, and G. Dudek, "Sketch Interpretation and Refinement Using Statistical Models", Eurographics Symposium on Rendering, 2004, pp. 23-32.

[3] D. Anderson, C. Bailey, and M. Skubic, "Hidden Markov Model Symbol Recognition for Sketch-Based Interfaces", AAAI Fall Symposium, 2004, pp. 15-21.

[4] H. Hse, and A. R. Newton, "Sketched Symbol Recognition using Zernike Moments", Pattern Recognition, Vol. 1, IEEE, Cambridge, UK, 2004, pp. 367-370.
[5] R. Arandjelović, and T. M. Sezgin, "Sketch Recognition by fusion of temporal and image-based features", Pattern Recognition, 44(6), 2011, pp. 1225-1234.
[6] T. M. Sezgin, and R. Davis, "HMM-Based Efficient Sketch Recognition", the 10th International Conference on Intelligent User Interface, ACM, 2005, pp. 281-283.
[7] T. M. Sezgin, and R. Davis, "Temporal Sketch Recognition in Interspersed Drawings", the 4th Eurographics Workshop on Sketch-Based Interfaces and Modeling, ACM, 2007, pp. 15-22.
[8] S. F. Qin, D. K. Wright, and I. N. Jordanov, "A conceptual design tool: a sketch and fuzzy logic based system", the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 215(1), 2001, pp. 111-116.
[9] W. Deng et al., "On-line sketch recognition using direction feature", IFIP Conference on Human-Computer Interaction, Springer Berlin Heidelberg, 2013, pp. 259-266.
[10] W. Shuxia et al., "Segmentation of Online Sketching Using Velocity Features", Journal of Northwestern Polytechnical University, 2, 2016, p. 8.
[11] B. Agrell, and O. Dehlin, "The Clock-Drawing Test", Age and Ageing, 27(3), 1998, pp. 399-404.
[12] S. Maechima et al., "Usefulness of a Cube-Copying Test in Outpatients with Dementia", Brain Injury, 18(9), 2004, pp. 889-898.
[13] K. K. F. Tsoi, N. W. Y. Leung, and M. W. Y. Lam, "An automated electronic rating system for the overlapped pentagons in dementia screening tests", Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 12(7), 2016, pp. 687-688.
[14] M. Mitolo et al., "The new Qualitative Scoring MMSE Pentagon Test (QSPT) as a valid screening tool between autopsy-confirmed dementia with Lewy bodies and Alzheimer's disease", Journal of Alzheimer's Disease, 39(4), 2014, pp. 823-832.
[15] M. F. Folstein, S. E. Folstein, and P. R. McHugh, ""Mini-mental" state: a practical method for grading the cognitive state of patients for the clinician", Journal of Psychiatric Research, 12(3), 1975, pp. 189-198.
[16] P. Caffarra et al., "The qualitative scoring MMSE pentagon test (QSPT): a new method for differentiating dementia with Lewy body from Alzheimer's disease" Behavioural Neurology, 27(2), 2013, pp. 213-220.
[17] R. H. Shumway, and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm", Journal of Time Series Analysis, 3(4), 1982, pp. 253-264.
[18] W. K. Ching et al., "Higher-Order Markov Chains", Markov Chains, Springer, US, 2013, pp. 141-176.
[19] A. E. Raftery, "A model for high-order Markov chains", Journal of the Royal Statistical, Series B (Methodological), 1985, pp. 528-539.
[20] Z. Ghahramani, and S. Roweis, "Learning nonlinear dynamical systems using an EM algorithm", In M. J. Kearns, S. A. S. and Cohn, D. A., editors, Advances in Neural Information Processing Systems, 11, 1999.
[21] C. A. Micchelli, Y. Xu, and H. Zhang, "Universal kernels" Journal of Machine Learning Research, 7(Dec), 2006, pp. 2651-2667.
[22] C. E. Rasmussen, and C. K. I. Williams, "Gaussian Processes for Machine Learning", MIT Press, 2(3), 2006.
[23] M. Lázaro-Gredilla et al., "Sparse Spectrum Gaussian Process Regression", Journal of Machine Learning, 11(2010), pp. 1865-1881.
[24] Z. S. Nasreddine et al., "The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment", Journal of the American Geriatrics Society, 53(4), 2005, pp. 695-699.
[25] D. Duvenaud, "Automatic Model Construction with Gaussian Processes", Doctoral Dissertation, University of Cambridge, 2014.
[26] R. M. Neal, "Bayesian learning for neural networks", Springer Science & Business Media, 118, 2012.
[27] A. Rahimi, and B. Recht, "Random Features for Large-Scale Kernel Machine", NIPS, 3(4), 2007, p.5.
[28] N. Wiener, "Generalized Harmonic Analysis", Acta matematrica, 55(1), 1930, pp. 117-258.
[29] A. Khintchine, "Korrelationstheorie der Station-ren Stochastischen Prozesse", Mathematische Annalen, 109(1), 1934, pp. 604-615.
[30] S. Bochner, "Monotone Funktionen", Stieltjessche Integrale und harmonische Analyse, Mathematische Annalen, 108(1), 1933, pp. 378-410.
[31] C. P. Robert, "Monte carlo methods", John Wiley & Sons, Ltd., 2004.
[32] R. M. Crum et al., "Population-based norms for the Mini-Mental State Examintation", JAMA, 269(18), 1993, pp. 2386-2391.
[33] J. E. Graham et al., "Prevalence and severity of cognitive impairment with and without dementia in an elderly population", The Lancet, 349(9068), 1997, pp. 1793-1796.
[34] R. G. Logsdon et al., "Assessing quality of life in older adults with cognitive impairment", Psychosomatic Medicine, 64(3), 2002, pp. 510-519.
[35] S. Artero et al., "Risk profiles for mild cognitive impairment and progression to dementia are gender specific", Journal of Neurology Neurosurgery & Psychiatry, 79(9), 2008, pp. 979-984.
[36] J. O. Berger, "Statistical decision theory and Bayesian analysis", Springer Science & Business Media, 2013.
[37] M. P. Petticrew et al., "False-negative results in screening programmes: systematic review of impact and implications", Health technology assessment (Winchester, England), 4(5), 1999, pp. 1-120.
[38] C. Andrade, and R. Radhakrishnan. "The prevention and treatment of cognitive decline and dementia: An overview of recent research on experimental treatments", Indian Journal of Psychiatry, 51(1), 2009, p. 12.
[39] J. W. Ashford, "Should older adults be screened for dementia?", Alzheimer's & Dementia, 2(2), 2006, pp. 76-85.
[40] D. Grossi et al., "The selective inability to draw horizontal lines: A peculiar constructional disorder", Journal of Neurology, Neurosurgery & Psychiatry, 64(6), 1998, pp. 795-798.

[41] J. G. Phillips et al., “Characteristics of handwriting of patients with Huntington’s disease”, Movement Disorders, 9(5), 1994, pp. 521-530.
[42] J. G. Phillips, G. E. Stelmach, and N. Teasdale, “What can indices of handwriting quality tell us about Parkinsonian handwriting?”, Human Movement Science, 10(2), 1991, pp. 301-314.
[43] R. L. Albin, A. B. Young, and J. B. Penney, “The functional anatomy of basal ganglia disorders”, Trends in neurosciences, 12(10), 1989, pp. 366-275.
[44] P. D. Thompson et al., “The coexistence of bradykinesia and chorea in Huntington’s disease and its implications for theories of basal ganglia control of movement”, Brain, 111(2), 1988, pp. 223-244.
[45] D. L. Jones et al., “Programming of single movements in Parkinson’s disease: comparison with Huntington’s disease”, Journal of Clinical and Experimental Neuropsychology, 14(5), 1992, pp. 762-772.
[46] M. J. Slavin et al., “Consistency of handwriting movements in dementia of the Alzheimer’s type: A comparison with Huntington’s and Parkinson’s diseases”, Journal of the International Neuropsychological Society, 5(1), 1999, pp. 20-25.
[47] K. K. Gnanalingham et al., “Motor and cognitive function in Lewy body dementia: comparison with Alzheimer’s and Parkinson’s diseases”, Journal of Neurology, Neurosurgery & Psychiatry, 62(3), 1997, pp. 243-252.
[48] R. A. Schmidt, “The search for invariance in skilled movement behavior”, Research Quarterly for Exercise and Sport, 56, 1984, pp. 188-200.
[49] J. R. Tresilian, G. E. Stelmach, and C. H. Adler, “Stability of reach-to-grasp movement patterns in Parkinson’s disease”, Brain, 120(11), 1997, pp. 2093-2111.
[50] H. Hefter et al., “Impairment of rapid movement in Huntington’s disease”, Brain, 110(3), 1987, pp. 585-612.
[51] M. F. Mendez et al., “Disorders of the visual system in Alzheimer’s disease”, Journal of Neuro-Ophthalmology, 10(1), 1990, pp. 62-69.
[52] M. F. Mendez et al., “Complex visual disturbances in Alzheimer’s disease”, Neurology, 40(3), 1990, pp. 439-439.
[53] L. Hansen et al., “The Lewy body variant of Alzheimer’s disease: A clinical and pathologic entity”, Neurology, 40(1), 1990, pp. 1-8.
[54] D. Galasko et al., “Clinical and neuropathological findings in Lewy body dementias”, Brain and Cognition, 31(2), 1996, pp. 166-175.
[55] A. Wong et al., “The validity, reliability and clinical utlity of the Hong Kong Montreal Cognitive Assessment (HK-MoCA) in patients with cerebral small vessel disease”, Dementia and Geriatric Cognitive Disorders, 28(1), 2009, pp. 81-87.
[56] R. Kummar, and A. Indrayan, “Receiver operating characteristic (ROC) curve for medical researchers”, Indian Pediatrics, Vol. 48, No. 4, 2011, pp. 277-287.
[57] J. A. Hanley, and B. J. McNeil, “The meaning and use of the area under a receiver operating characteristic (ROC) curve”, Radiology, 143(1), 1982, pp. 29-36.
[58] P. S. Sachdev et al., “The prevalence of mild cognitive impairment in diverse geographical and ethnocultural regions: the COSMIC collaboration”, PloS one, Vol. 10, No. 11, 2015.