Modeling and Analytics of Multi-factor Disease Evolutionary Process by Fusing Petri Nets and Machine Learning Methods

Wangyang Yu[a,b], Xuyue Wang[a,b,*], Xianwen Fang[c] and Xiaojun Zhai[d]

[a]Key Laboratory of Intelligent Computing and Service Technology for Folk Song, Ministry of Culture and Tourism, Shaanxi Normal University, No. 620, West Chang'an Street, Chang'an District, Xi'an, 710119, China
[b] School of Computer Science, Shaanxi Normal University, No. 620, West Chang'an Street, Chang'an District, Xi'an, 710119, China
[c]Department of Computer Science and Engineering, Anhui University of Science and Technology, An'hui, 232001, China
[d]School of Computer Science and Electronic Engineering, University of Essex, Colchester, CO4 3SQ, U.K.

ARTICLE INFO

Keywords:
Evolutionary Process
Machine Learning
Data Analytics
Multifactorial Disease
Coloured Petri nets

ABSTRACT

Recent years, informatization methods have been gradually applied to medical treatment, in which machine learning and evolutionary computation play an important role. However, the effective methods for the study of multi-factor disease evolutionary process are still largely open. There are some issues in the field of disease analysis, such as the lack of visual multi-factor disease evolution model and effective analysis methods. For a universal method of data analysis and medical diagnosis, the machine learning algorithms should be combined with the formal modeling methods to fully realize the complementary advantages, make model has the advantages of visualization and efficient data analysis. This work proposes a novel research idea for the modeling analysis of current multi-factor diseases and reveal its feasibility, so as to explore potential pharmaceutical targets and enable doctors and patients to better understand the evolution process of multi-factor diseases. It is worth mentioning that, in order to verify the feasibility of the proposed idea, we applied it to the analysis of the role of monoamine hormones in depression. The model incorporates the machine learning algorithms, and it finally outputs the pathogenic probability under different hormone levels, reflecting the importance of different factors on depression. The application case proved that we provide a clear process model and a novel research method for multi-factor disease evolutionary process analysis.

1. Introduction

The advancement of Internet technology has led to an increase in the application of research methods in the medical field [1, 2, 3, 4, 5]. However, the complexity of disease development is significant. The interaction of external factors and the body's reaction functions can result in changes in the structure, metabolism, and function of the body. Understanding these changes is essential for studying and preventing diseases [6]. The field of pathology is critical in identifying the causes, mechanisms, and rules of disease development [7, 8, 9, 10]. In addition, research on multi-factor disease analysis has been conducted, but currently, there is a lack of a suitable method for visualizing the evolutionary process of diseases, specifically in terms of displaying the impact of disease and the probability of different factors. To address this, we propose a method that combines formal methods and machine learning algorithms for multi-factor disease evolutionary process analysis. This approach aims to resolve the non-visual aspect of disease progression and aid doctors in analyzing and observing the effects of various factors on disease development.

In this paper, in order to facilitate the model construction, we firstly proposed a generalized model of multi-factor disease analysis using Coloured Petri nets (CPN) [11], and then used the refined the original generalized model through CPN for specific diseases. The advantages and progressiveness of this model are mainly reflected in the visualization of the process, and innovatively incorporating machine learning algorithms to analyze the data. In the specific diseases, CPN is used to model the action evolutionary process of multi-factor diseases and analyze the action details of each factor. As an effective modeling tool, CPN is a kind of discrete event modeling language, and it is a combination of Petri nets [12] and high-level programming languages. At the same time, CPN can dynamically reflect the running states and the relationships among various parts. In CPN, different factors are characterized, and the machine learning

*Corresponding author
ywy191@snnu.edu.cn (W. Yu); wangxuyue@snnu.edu.cn (X. Wang); fang_xw@outlook.com (X. Fang); xzhai@essex.ac.uk (X. Zhai)
ORCID(s): 0000-0002-2102-4410 (X. Wang)

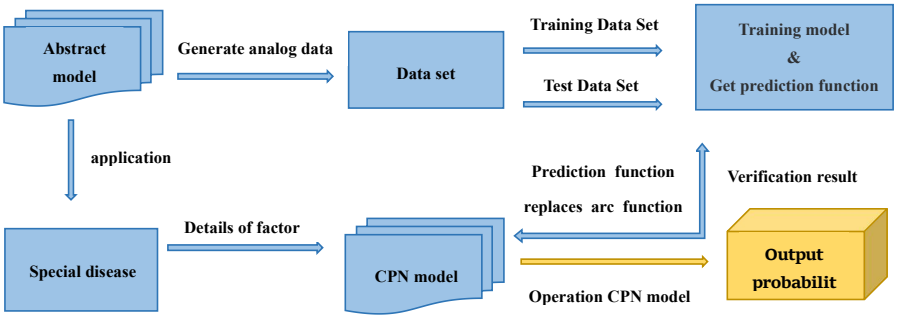

Figure 1: The proposed methodology to model and analyze multi-factor disease processes

algorithms is used to process the data and output the probability of a disease, so as to reflect the influence degree of different factors in the disease.

According to the statistics of the World Health Organization, the global incidence of depression is about 11%, and the number of patients is as high as 340 million [13]. Nowadays, with the increasing pressure of contemporary young people, the incidence of depression is increasing significantly around the world, and the depression has become the fourth largest disease in the world. Therefore, we chose depression as our main target diseases in our research.

As of now, most studies are limited to the state manifestation of depression patients, and its pathological role is not clear in medicine, among the relationship between various monoamine hormones also cannot be clearly presented in graph view [14]. Thus, it is particularly important to pay attention to the pathological roles of depression-related hormones in the treatment of depression, we are necessary to construct a model to depict and analyze these associations. In complex evolutionary systems such as multi-factor diseases, visual modeling can show more pathological details. So we introduce visual modeling into the description of monoamine hormones. At present, most doctors rely on QAs to diagnosing depression. In contrast, the advantage of our modeling approach is that it follows data analysis and reduces the subjective judgment of doctors. In the case study of depression, the model can analyze specific states caused by three different levels of monoamine hormones, and output the disease probability of different factors in the depression. In the modeling process, we utilize hierarchical CPN to visualized the hormone effects process. The hierarchical CPN can simplify complex models, which is fit with the analysis process of diseases under different factors. The hierarchical CPN can divide each factors into different sub-modules, which make the structure of each part clearer. Through model integration and simulation, we come to the conclusion that low levels of the three monoamine hormones increased the risk of the disease. The main contributions of this paper are as follows, and the framework of the proposed method is shown in Fig. 1.

1) The framework: We propose a modeling and analytics framework of multi-factorial disease processes, which includes an analytical method that effectively combines CPN and machine learning algorithms, and it is ultimately applied to study the evolutionary process of multiple monoamine hormones in depression.

2) The modeling approach: We first proposed a generalized model for multi-factor symptom analysis based on CPN, which is suitable for analyzing most multi-factor diseases, and then applied it to a specific case. The core significance is to propose a novel modeling approach for the multi-factorial diseases process analysis.

3) The analytical method: In order to better analyze the final impact of each factor on the pathogenic results, we use the machine learning algorithms to analyze the data, and replace the arc functions to output the probability of disease occurrence, so as to reflect the influence degree of different factors on the disease. This allows us to observe the probability of the disease more intuitively, under different facts.

The rest of this paper is organized as follows: The second section introduces the related work. The third section describes the whole construction process of multi-factor disease analysis model. The fourth section describes the specific analysis methods and simulation results of the model. The fifth section summarizes the full paper.

2. Related Work

In recent years, many healthcare systems have used various types of Petri nets as mathematical tools to create models and analyze them [15]. In the latest studies, the authors proposed an infectious disease transmission model based on Generalized Stochastic Petri nets, which qualitatively analyzes the transmission modes of COVID-19 and other infectious diseases at various stages [16]. [17] and [18] establish an extended model to predict independent emotional components based on CPN. In [19], using CPN, the authors proposed a modeling method to visualize machine learning algorithm prediction processes. In [20], the paper proposed a Fuzzy-stochastic Mixed Petri nets as a mathematical tool to model the complex psychological system, and established an emotion regulation model. In [21], Petri nets was used to establish a relationship model between mood and sleep time. The Mani Mehraei *et al*. [22] analyzed the simulation results mainly by manipulating the processing rate of the hybrid Petri nets model. For multi-factorial analysis of disease evolution, Ref. [8] proposed a method that confirmatory factor analysis of single versus multi-factor models for Alzheimer's disease. Ref. [9] proposed a strategy that combined Multi-factor biomarker screening to rapidly diagnose Alzheimer's disease. Ref. [10] summarized multiple aspects of cancer evolution. These studies give us some enlightenments. However, they are not enough to cope with the pathogenesis of multi-factorial diseases.

Machine learning is widely used in medical image analysis and disease prediction, as well as other relevant fields [23, 24, 25]. Due to the deviation of the algorithm, the prediction results are still susceptible to the influence of concept drift, inexplicability, label bias and errors [26, 27, 28]. Formal methods can be used to solve such problems. Among many formal models, Petri nets [29, 30, 31] is a powerful modeling tool. It can accurately describe the process behavior, and have interpretability. These advantages make Petri nets to some extent make up for the shortcomings of machine learning and ensure the accuracy of machine learning results [32].

On the basis of above research, in order to retain the advantages of formal language modeling and take into account the advantages of machine learning in data analysis, we propose a hybrid multi-factor disease analysis model, which combined machine learning algorithms and Petri nets to model and analyze multi-factorial diseases evolutionary process. It can analyze the interactions among multiple factors. The model includes univariate and multi-factorial processes. Since we want to incorporate machine learning, we adopt CPN that a high level Petri nets. This is because in the data analysis part, we use the machine learning algorithm to train the related data to obtain the prediction functions, and replace the traditional arc functions of CPN with them. We need to characterize different factors as different color sets to give them a special biological meaning, so CPN is used in the proposed methodology. CPN regards different action processes as sub-modules, which makes the structure clearer. In the model simulation, we integrated machine learning algorithms to process relevant data, and finally outputted the impact ratio of various factors. Then, we conducted in-depth research into the mechanism of depression, as which a multi-factor disease analysis application example of the proposed method. In the case study, we use CPN to describe the depression of interaction among a variety of single amine hormones, and construct the relationship model among dopamine, serotonin and norepinephrine. In summary, our method not only well visualizes the entire evolutionary process of multi-factor disease, but also conducts simulation and outputs relevant results, to help doctors develop new strategies for the treatment of psychological disorders.

3. Modeling Scheme

3.1. Related Concepts

3.1.1. CPN Module and Hierarchical CPN

Petri nets is traditionally divided into low level Petri nets and high level Petri nets. CPN belongs to the high level Petri nets, which is an aggregate that can combine Petri nets with programming languages, and reduces the basic elements in the network system through the classification of tokens, so as to achieve the purpose of reducing the scale of the network system. Generally, the goal of model simulation is to debug and study system design [33]. We can studied the behaviors of the system in different situations through the simulation of CPN model, so CPN model is simpler and clearer when modeling complex system. Based on these advantages of CPN, we establish an abstract model based CPN, and applied it to the multi-factor study of depression. The definition of CPN can refer to [11], and the following is the definitions of the CPN Module and the hierarchical CPN.

Definition 1 [11]: A *CPN Module* is a four-tuple, $CPN_M = (CPN, T_{sub}, P_{port}, PT)$, where:

1) $CPN = (P, T, A, \Sigma, V, C, G, E, I)$ is a non-hierarchical Coloured Petri Nets.

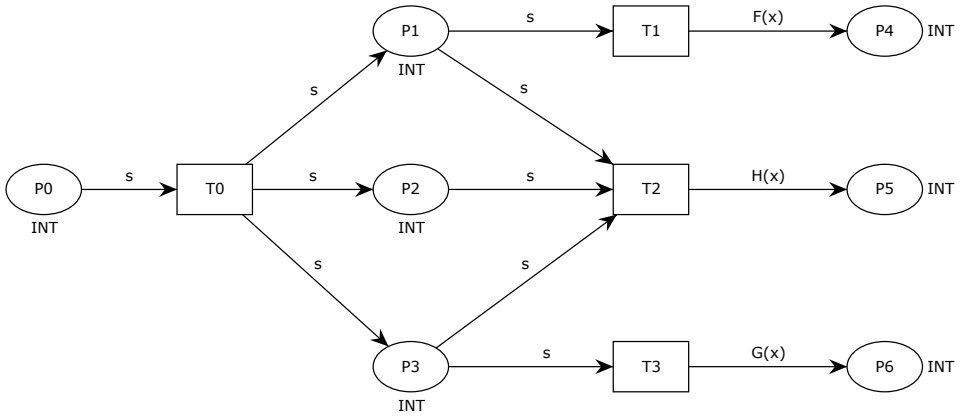

Figure 2: The MDP model

Table 1
The meaning of places

Place	Meaning
P_0	Disease factor input
P_1	Factor 1
P_2	Factor 2
P_3	Factor 3
P_4	Probability 1
P_5	Probability 2
P_6	Probability 3

2) $T_{sub} \subseteq T$ is a set of substitution transitions.

3) $P_{port} \subseteq P$ is a set of port places.

4) $PT : P_{port} \rightarrow \{IN, OUT, I/O\}$ is a port type function that assigns a port type to each port place.

Definition 2 [11]: A *hierarchical CPN* is a four-tuple $CPN_H = (S, SM, PS, FS)$, where:

1) S is a finite set of modules. Each module is a CPN Module $s = ((P^s, T^s, A^s, \Sigma^s, V^s, C^s, G^s, E^s, I^s), T^s_{sub}, P^s_{port}, PT^s)$. It is required that $(P^{s_1} \cup T^{s_1}) \cap (P^{s_2} \cup T^{s_2}) = \emptyset$ for all $s_1, s_2 \in S$ such that $s_1 \neq s_2$.

2) SM: $T_{sub} \subseteq S$ is a sub-module function that assigns a sub-module to each substitution transition.

3) PS is a port-socket relation function that assigns a port-socket relation $PS(t) \subseteq P_{sock}(t) \times P^{SM(t)}_{port}$ to each substitution transition t. It is required that $ST(\mathrm{p}) = PT(p')$, $C(p) = C(p')$, and $I(p)<> = I(p')<>$ for all $(p, p') \in PS(t)$ and all $t \in T_{sub}$.

4) $FS \subseteq 2^P$ is a set of non-empty fusion sets such that $C(p) = C(p')$ and $I(p)<> = I(p')<>$ for all $p, p' \in fs$ and all $fs \in FS$.

3.1.2. Multi-factor disease analysis model

We proposed an abstract *Multi-factor Disease Process* (MDP) model, which is on the basis of CPN. The MDP model accomplish the hierarchy by hierarchical CPN, making the model more organized and the logic of disease influences clearer. The following is the definition of MDP.

Definition 3: A MDP is a hierarchical CPN = $(P, T, A, \Sigma, V, C, G, E, I)$ that satisfies the following three points:

1) A MDP has three kinds of special places α, $B = (\beta_1, \beta_2, \beta_3, \cdots, \beta_n)$ and $\Gamma = (\gamma_1, \gamma_2, \gamma_3, \cdots, \gamma_n)$, where $n \in \mathbf{N}^+$; $\alpha \in P$ is called the source place that represents the set of various disease factors; $B \subseteq P$ is called the sink place set that represents the output values of the each factor, and ${}^\bullet\{\alpha\} = B^\bullet = \emptyset$; $\Gamma \subseteq P$ is called branch place set that represents the influencing process of disease each factor.

2) A MDP has two kinds of special transitions δ and $Z = (\zeta_1, \zeta_2, \zeta_3, \cdots, \zeta_n)$, where $n \in \mathbf{N}^+$; $\delta \in T$ is called the classified transition that represents α is divided into specific processes at different levels of Γ; $Z \subseteq T$ is called the

Table 2
The meaning of transitions

Transition	Meaning
T_0	Data collection
T_1	The action process of factor 1
T_2	The action process of factor 2
T_3	The action process of factor 3

Table 3
The meaning of arc functions

Arc Function	Meaning
$F_{(x)}$	Predictive function of factor 1
$H_{(x)}$	Predictive function of factor 2
$G_{(x)}$	Predictive function of factor 3
s	Influencing factors in specific cases

hierarchical transition set that consists of a finite set S of modules, $s \in S$ is defined according to the Definition 2. ζ_1, ζ_2, ζ_3, $\cdots$, ζ_n can separately represent a specific response to one of the factors each module.

3) A MDP has one kind of special arcs $F = (f_1, f_2, f_3, \cdots, f_n)$, where $n \in \mathbf{N}^+$; $F \subseteq A$, $E(f_n) = Func(v_1, v_2, v_3, \cdots, v_n)$, *Func* is the prediction function derived from the relevant machine learning algorithms, $v_1, v_2, v_3, \cdots, v_n \in V$.

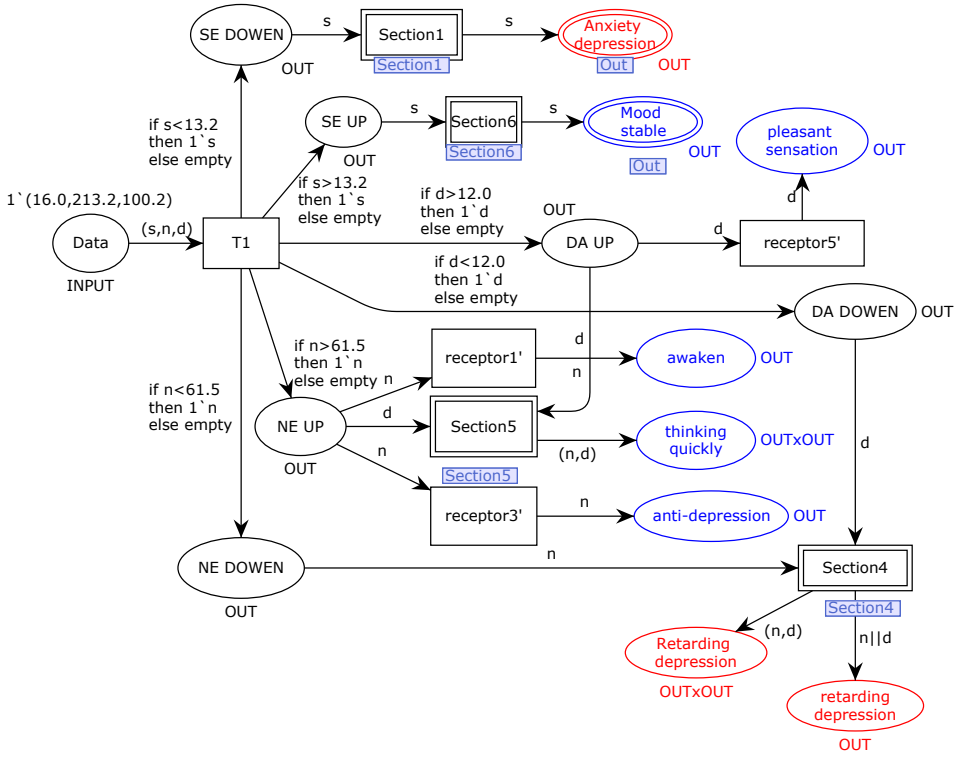

Figure 3: Top-level MDP module of depression

In MDP model of Fig. 2, P_0 is the input that is a disease factor, and after the judgment T_0, the next step will be the cross-action of various factors. The specific change process of each factor is abstracted by the transitions T_1, T_2 and

T_3, and the corresponding results are finally outputted by the arc functions $F(x)$, $H(x)$ and $G(x)$. Table. 1-3 shows the notations used in the data collection and analysis process in Fig. 2.

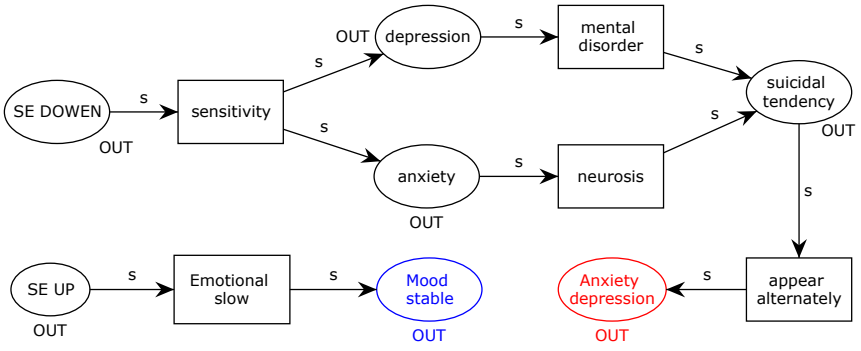

Figure 4: Part 1 and Part 2

3.2. The modeling process of case study

In this section, we applied the previously proposed MDP model into a specific example, establishing a MDP model about dopamine, serotonin and noradrenaline in depression mechanism. The model is constructed empirically from existing medical assertions and is also based on data-driven. This is due to the fact that the problem we study is illustrated based on data. Inspired by these backgrounds, we constructed a generalized research idea in multi-factorial diseases, and introduced a relevant dataset based on that. Driven by a combination of data and existing expert experience, our modeling ideas become more convincing. CPN provides the foundation and basic primitives for graphical representation. Its goal is to model specific systems with formal modeling methods. We need the abstract model at the top level to allow us to focus on the details of actions. CPN modules can be seen as unknown boxes. The core advantage of hierarchical CPN is that it allows model actions to work at different levels of abstraction without interfering with each other. Therefore, after establishing an abstract model by CPN, we need to consider multiple factors of disease, regard each factor as a sub-part of the model, and use machine learning algorithms to predict the pathogenic probability of various factors, and final achieve the hierarchical modeling.

Before modeling the entire mechanism process, we firstly need to consider the consequences that may be involved under different hormones. The next step is to judge the state under different hormone levels according to the input conditions, meanwhile, we give the allocation principles and the problems that should be paid attention during the modeling process, and obtain the formal model. Finally, We finish the hierarchical modeling of the whole system. Throughout the modeling process, we have done the following tasks: according to medical rules, we input a person's three hormone levels as a color set, in which the values of each element are given different data types. Then this input will be operated through various judgment rules. The rules of specific state and structure are extracted from clinical characterizations. Below is a detailed explanation of the steps.

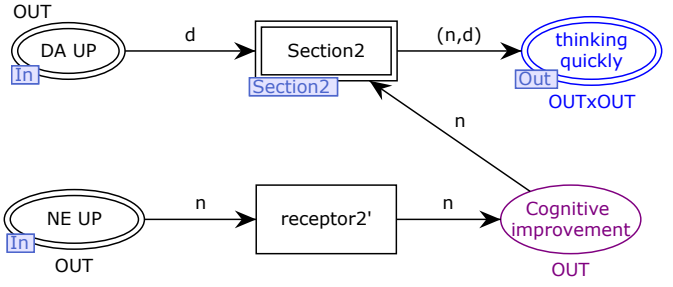

Figure 5: Part 3

3.2.1. Hormone judgment Principles

The action mechanism of hormone changes should follow the corresponding pathological features and medical standards [34, 35, 36]. Therefore, for the action mechanism of the above three hormones, we need to input the corresponding hormone levels. For serotonin, the normal range in human blood is 161.45 ± 31.3ng/ml, and the normal range of corresponding norepinephrine and dopamine in human body is 615 ± 3240pmol/ L and 225.76 ± 104.83g/24h. When hormone range is lower than the normal range, corresponding state changes will occur. In the simulation data, when the hormone value is within the normal range, set it's judgment value as 1, and vice versa to as 0. A simulated data set consists of dopamine, norepinephrine, serotonin, and a final decision value, which is made up of random numbers generated by Python within medical range. Based on this, we carried out the hierarchical modeling, which requires modeled the state of each hormone production, and then integrated them together.

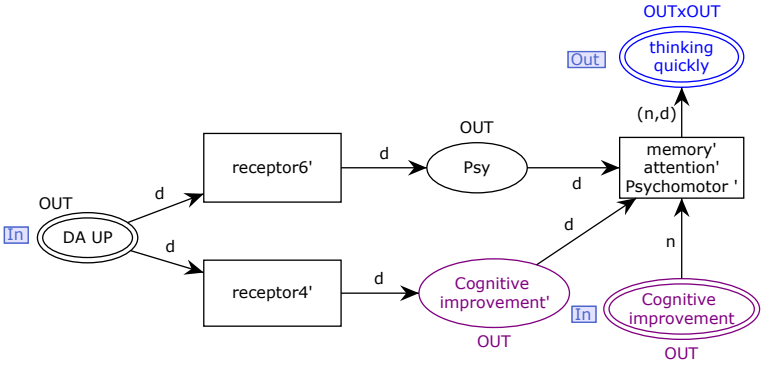

Figure 6: Part 6

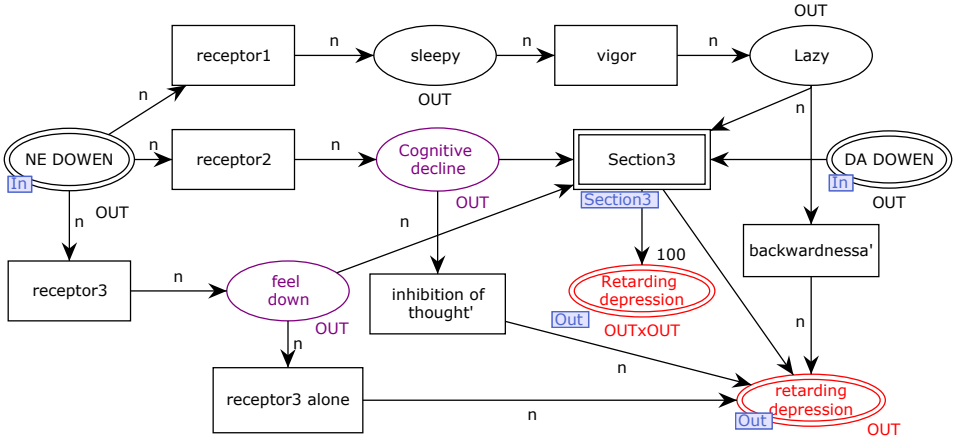

Figure 7: Part 4

After considering all the required occurrence conditions, we also need to pay attention to the state changes caused by single action, joint action, and concurrent state. In the entire model, our hormone behavior model is a formal modeling for a specific situation, so it also need to be adhered to the following guidelines:

1) Hormone action cannot be interrupted until reaching final state, and hormones don't have any priority order between evolutionary process.

2) In the CPN model, we combined the content of three hormones into a marked set as the input, and finally formed multiple marked sets of different hormones as the output. For example, the output of a single hormone is one marked set, and the output of multiple hormones is another marked set.

3.2.2. The complete model

Modeling the action mechanism of monoamine hormones involves three key steps. Many different hormones usually act at the same time, so we need to judge the hormone level. After the hormone value is input into T_1, we need to judge whether the hormone content is lower than the minimum level through arc expression, and then carry out state classification, which is shown in Fig. 3.

The color set in Fig. 3 is declared as follows:

closet OUT = real ;
*closet INPUT = product OUT * OUT * OUT* ;
*closet OUT x OUT = product OUT * OUT* ;
var s , n , d , z , m : OUT .

Fig. 3 is the top-level MDP module of depression, which includes two levels and six parts, and then we will show the details of the other two levels and six parts, as shown in Figs. 4-8. The top-level MDP module of depression represents a abstract view of the entire MDP module of depression. In the MDP module of depression, we can see that the *INPUT*, intermediate frame, and *OUT* exchange tokens with each other, via the places and transitions, but we cannot see the specific structure of what the *INPUT*, intermediate frame, and *OUT* do. In Fig. 3, 1'(16.0,21.3,10.2) represents that a person's levels of three hormones, under normal circumstances, the value should be ten times that of it. But in order to facilitate the subsequent machine learning algorithm to be integrated into the model for analysis, we uniformly reduced the input data to one-tenth.

In Fig. 4, Part 1 and Part 2 respectively describes a series of physical changes and subsequent results caused by normal and low levels of serotonin in human body. The increase of serotonin will make the mood gradually tend to a stable state; on the contrary, it's lower can lead to emotional instability, increase the risk of depression, and eventually lead to anxiety depression. In the analysis process, we will integrate the machine learning algorithms into the specific action mechanism. The detailed combination methods are described in Section 4.

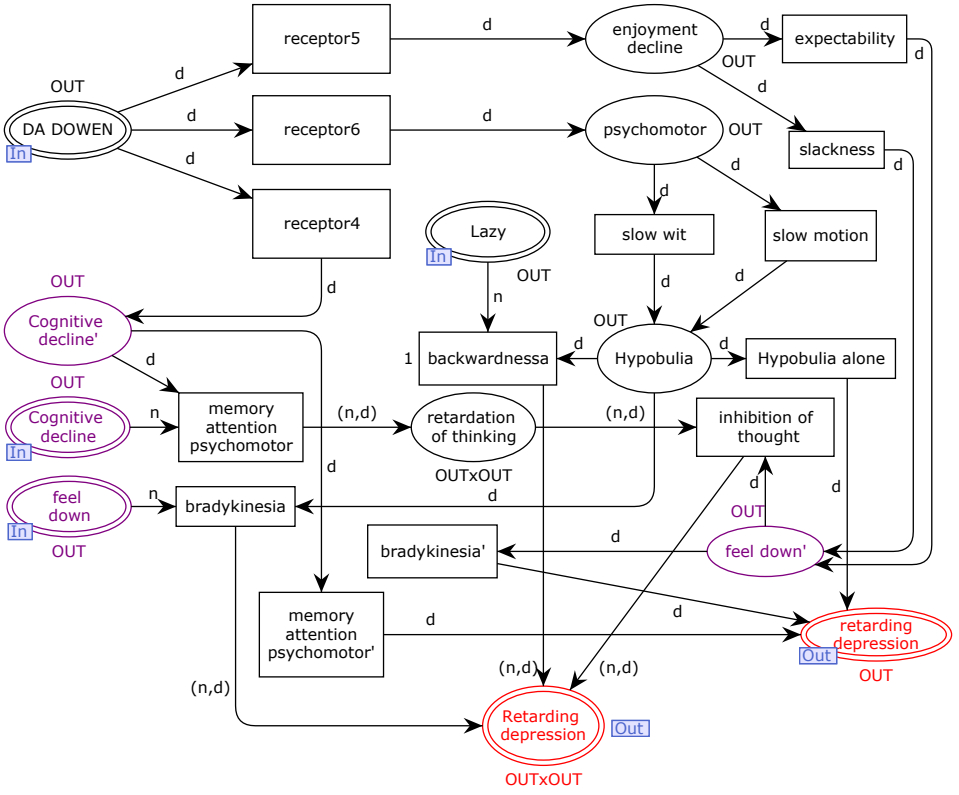

Figure 8: Part 5

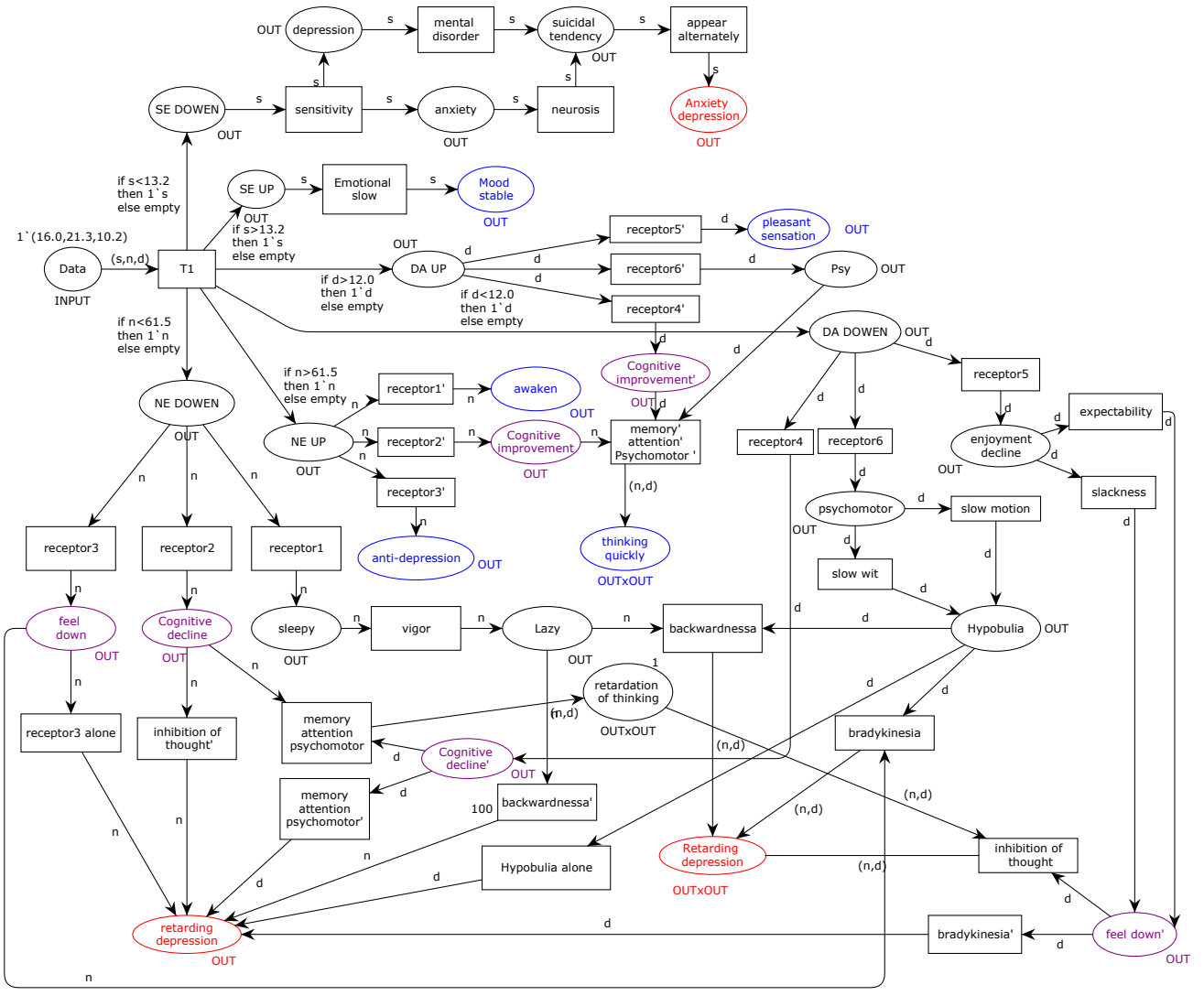

Figure 9: The complete model

Since low norepinephrine and dopamine both increased the risk of retarding depression, we can combine the both in modeling to represent their interaction, as shown in Figs. 5-8. Fig. 5 and Fig. 6 show Part 3 at the second level and Part 6 at the third level, we can see that Part 3 contains Part 6, and Part 6 reflects the interaction between norepinephrine and dopamine content in the normal range, so there are two variables n and d in the last transition, and the final output type is $OUTxOUT$. By the mechanism of action, it can be known that when its content increases, it can reduce the probability of disease.

Finally, Fig. 7 and Fig. 8 respectively shows Part 4 for the second level and Part 5 for the third level. From Fig. 7, we can see that Part 5 is a composition of Part 4, and the two parts jointly reflect the common connection between the low levels of norepinephrine and dopamine. The decline of norepinephrine will resulted in inadequate α_1 and α_2 receptors, while a lack of dopamine results in inadequate D_1 receptors in the prefrontal cortex, D_1 receptors in the nucleus accumbens and D_2 receptors in the striatum. Although the two hormones act on different receptors, they lead to the same state, so they are interacted. It can be seen from Fig. 8 that the output type of single variables n and d is OUT, which represents the result after the action of single hormone. However, when the output type is $OUTxOUT$, there are multiple variables n and d simultaneously exist in a transition, which represents the result of the interaction of various hormones.

Fig. 9 shows the details of the complete model. By the hierarchical modeling, the model mechanism becomes simpler and easier to analyze and understand. At the beginning of this section, we show the relationships among different levels of hormones and different states. At here we show the details of the effects of different levels of hormones. In the whole model, the balance of the action mechanism among the three hormones can be clearly demonstrated. We will verify the rationality of the model through model simulation in the next section.

4. Model simulation and analysis

In this section, we integrate *Logistic Regression algorithm* into MDP model of depression. In accordance with the relevant medical standards, we generate some simulated data, and training data, obtained prediction function and forecasting results. Finally, we replaced the arc functions with the prediction functions to get the result of simulation, and verified the rationality and validity of the model. The specific process of model simulation is described in detail below.

4.1. Algorithm Analysis

We select the *Logistic Regression algorithm* [37, 38] among many machine learning algorithms for the disease classification. The reason why we chose the logical regression algorithm is that the parameters of its expression conform to the expression structures of the CPN arc functions. That makes the integration of CPN and machine learning more convenient. The results of the logical regression algorithm are between 0 and 1, which conforms to the result setting of the polarization of most diseases. For the research problem of multi-factor disease analysis evolutionary process, we need to consider the global situation rather than individual differences. The hypothesis of the *Logistic Regression algorithm* model is $h(\Theta)= g(\Theta^T X)$, where: X represents the feature vector, and g stands for the logical function. The algorithm is a commonly used Sigmoid function, and its formula is $S(x) = \frac{1}{1+e^{-x}}$. The function image is shown in Fig. 10.

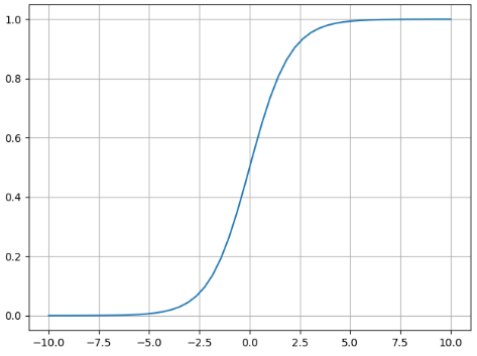

Figure 10: Sigmoid function image

For a given input variables, according to the selected parameters, $h_\Theta(x)$ can calculate an estimated possibility that output variable is equal to 1 : $h_\Theta(x) = P(\ y = 1 \mid x\ ;\ \Theta\ )$. If we calculates $h_\Theta(x) = 0.7$, this means that the disease has 70% possibility being a positive type. Correspondingly, if it equals 0.3, this means that has 30% possibility of being a negative type. Before the algorithm analysis, we generate some simulation data according to relevant medical guidelines for our train and test. The data is mainly composed of *SE*, *NE*, *DA* and *Output* . The value of *Output* is 0 or 1, which represents whether a person is depressed in a given situation. It is mainly determined by the values of *SE*, *NE*, and *DA*.

We analyze the data using *Logistic Regression algorithm*, which is implemented by Python. The method obtains the training model by the training of sample data, and then uses the training model to predict unknown data. In MDP model of depression, according to the normal levels of dopamine, norepinephrine, and serotonin in medical, we set three thresholds of 132, 615, and 120.93. Similarly, to facilitate the integration of machine learning algorithms into the model for analysis, we also reduced it to one-tenth of the normal data. For serotonin, when $h_\Theta(x)$ is less than 13.2, then $y = 1$; when $h_\Theta(x)$ is greater or equal to 13.2, then $y = 0$, and so on. In the process of modeling and evaluation, since the dictionary is unordered, we introduced the *OrderedDict* function to make the dictionary ordered, and converted

the data set into 2D data structure. So as to facilitate the extraction of features and labels by the *Pandas* function. For cross-validation, we randomly select data from the training and test samples at a ratio of 7 : 3. The model is trained with training data by import *fit* function into training model, and then the model is evaluated with test data. We use *exp* function in *Numpy* to obtain the probability value. Finally, we get the model predicted results of three hormone single action and interaction.

4.2. Model Simulation

As can be known from the above, we have obtained the four prediction functions generated by the single and interaction model. In the last step, we need to combine the obtained prediction functions with MDP model of depression. We replaced the output arc functions with the prediction functions to get the corresponding probability. The following are the specific steps of the model simulation.

In the final output arc functions, we used CPN own structures to achieve the loop structure in the algorithm, and inputted the relevant parameters, which are obtained from the *Logistic Regression* model. The correlation algorithm of the final model is shown in Algorithm 1. For Algorithm 1, it mainly represents the related rules of model operation. The model starts to run with the initial marking, its structure includes a loop, and this can show all possible transition sequences. Only when the former set of the current transition meets the trigger conditions, there will be token flow to the subsequent. In Algorithm 1, the arc function E is the function obtained by machine learning, and t_i is the probability of the output. Algorithm 1 can give a better understanding of how CPN works and how to get results. In the whole operation process of the model, the sub-modules of each hormone will also run according to Algorithm 1 and integrate into large modules through hierarchical CPN. Then, we used *CPNTools* [39] for model simulation. *CPNTools* is a

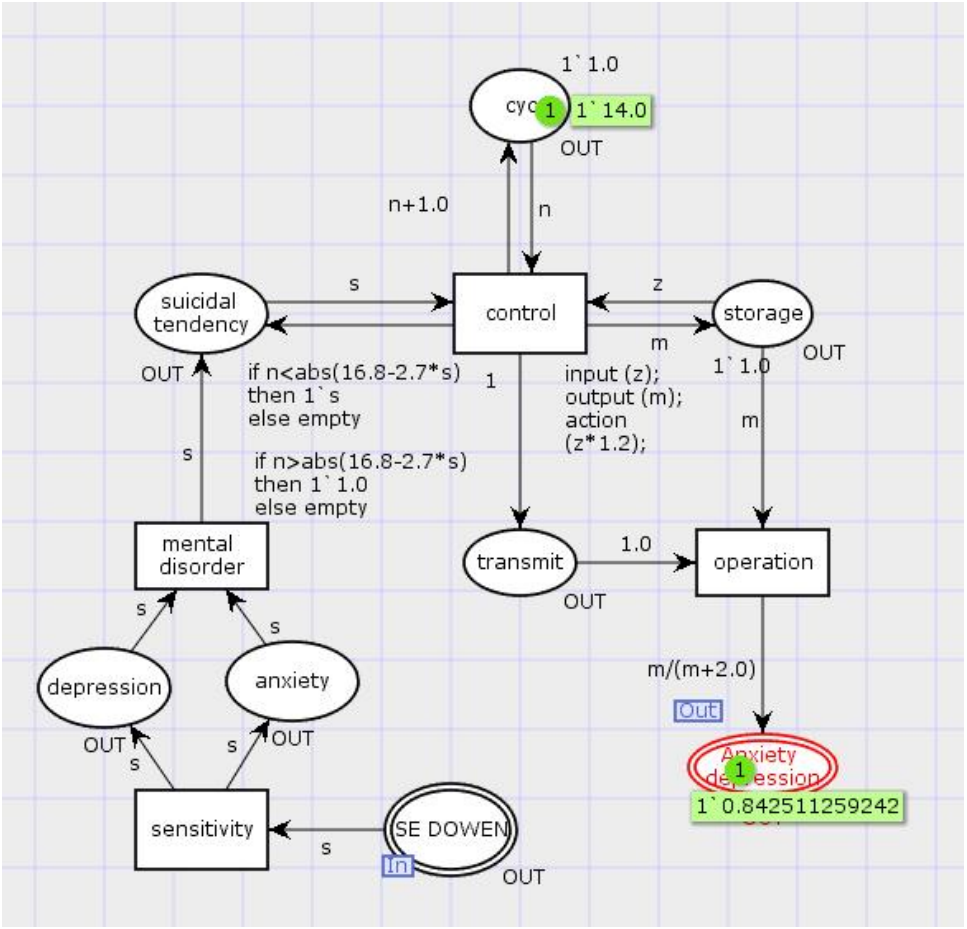

Figure 11: The model simulation result of SE down

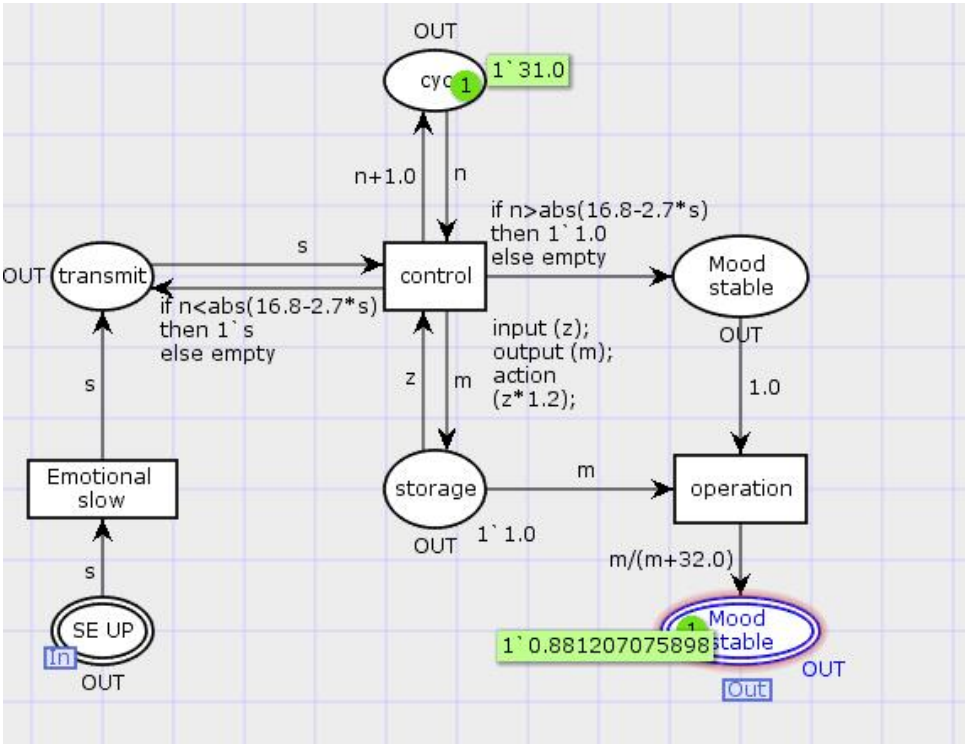

Figure 12: The model simulation result of SE up

Algorithm 1: Analysis of multiple pathogenic factors

Input: $CPN = (P, T, F, \Sigma, C, G, E, I), \langle M_0, Q, W \rangle$
Output: $CPN_2 = (P_2, T_2, F_2, \Sigma_2, C_2, G_2, E_2, I_2)$
$P_2 = \{p_i | M_0(p_i) \neq \emptyset\}$, $T_2 = \emptyset$, $F_2 = \emptyset$, $\Sigma_2 = \{C(p_i) | p_i \in P_2\}$, $C_2 = C(p_i)$, $G_2 = \emptyset$, $E_2 = \emptyset$, $I_2 = I(p_i)(p_i \in P_2)$.
Compute all enabled binding elements $\langle t_i, b_j \rangle$ under initial marking M_0, then get all enabled transitions t_i;
while $\{t_i | i = 1, 2, ..., n\} \neq \emptyset$ **do**
  if *Tokens.value satisfies the criteria* **then**
    for *each tokens in model* **do**
      if $p_i = \cdot t_i$ **then**
        $p_{i+1} \leftarrow token$
        $T_2 = T_2 \bigcup \{t_i\}$, $E_2 = E_2 \bigcup \{E(p_i, t_i)\}$, $A_2 = A_2 \bigcup \{(p_i, t_i)\}$, $G_2 = G_2 \bigcup \{G(t_i)\}$
      else
        $T_2 = T_2 \bigcup \{t_i\}$, $A_2 = A_2 \bigcup \{(p_i, t_i)\}$, $E_2 = E_2 \bigcup \{E(p_i, t_i)\}$, $G_2 = G_2 \bigcup \{G(t_i)\}$
        for *each* $p_i = t_i \cdot$ **do**
          $p_{i+1} \leftarrow token$
          $P_2 = P_2 \bigcup \{p_i\}$, $A_2 = A_2 \bigcup \{(t_i, p_i)\}$, $E_2 = E_2 \bigcup \{E(t_i, p_i)\}$, $C_2 = C_2 \bigcup \{C(p_i)\}$, $\Sigma_2 = \Sigma_2 \bigcup \{C(p_i)\}$
      E = Function obtained by the **Logistic Regression** algorithm
  Get the new marking M, recalculate the enabled binding elements $\langle t_i, b_j \rangle$ under the marking, where t_i is printed with probability

convenient simulation tool, which can dynamically display the action process and hormones change of model, and output the final states. At here we only show the final output models of serotonin, the simulation results as shown in Fig. 11 and Fig. 12.

The innovation of our method is reflected in the introduction of a novel research idea in multi-factor disease evolutionary process analysis. This new research idea can visualize the specific changes of different influencing factors, and analyze the data of different individuals, thus achieving the coexistence of personalization and visualization. Visual modeling is a way of thinking about problems; it can be modeled around realistic ideas. In this study, visual modeling is specifically applied to show the details of each depression symptom, and also to run the system dynamically during the simulation phase of the model, allowing dynamic tracking of each step of model changes. The advantage of visual

modeling is that it allows retaining the necessary details of complex problems and filtering out the unnecessary ones. It also provides a mechanism to observe the special system from different perspectives. The combination of CPN and machine learning makes maximum use of each other's advantages and makes up for each other's disadvantages. At the same time, it also overcomes the shortcomings of CPN in data analysis. In the whole process of fusion between CPN and machine learning, we bring the obtained prediction functions parameters directly into the CPN arc functions. These prediction functions are analyzed by data, and these analog data are set in accordance with relevant medical guidelines, so the fusion process and steps are strict and clear.

4.3. Discussion

In this part, we will discuss and verify the results. Fig. 11 and Fig. 12 shows the prediction process of the MDP model in depression. In Fig. 11, we can see that the input value for the serotonin hormone is 11.0, below normal range, and the probability tending to negative of output prediction is 0.842. In Fig. 12, When the input value of serotonin hormone is 17.0, within the normal range, the probability tending to positive of output prediction is 0.881, which is consistent with the results obtained by the training model in *Logistic Regression* model. To sum up, the simulation results can verify the rationality of model. Petri nets has the specific mathematical analysis method, it can view problems at a more abstract level, and also provides mature graphic expression. In addition, Petri nets can depict the system evolutionary process and provide a visual model to trace behavior changes, so as to analyze the causality of pathogenic factors and the mechanism of disease evolution and diffusion. However, Petri nets lack enough data analysis ability when facing multi-factor disease evolutionary process. For machine learning, its advantage lies in the data analysis ability, but there are also problems such as lack of visual modeling capability, label bias and errors. The effective combination of the Petri nets and machine learning can make up for their shortcomings, and make full use of their respective advantages. The above is the advantages of our modeling method. Of course, in the existing researches, this is the first attempt to use a formal approach to the analysis of multi-factorial disorders and to apply it to depression. Meanwhile, MDP has certainly universality, this is another advantage of this method. Although our existing work needs to be improved, especially in the fusion of CPN model and machine learning algorithm, this method still provides new ideas for research in the field of multivariate disease analysis.

Next, we will use more abstract formal methods to prove the relationship among influencing factors in various diseases, and explore more effective ways to combine CPN with machine learning. The advantages and progressiveness of this model are mainly reflected in the visualization of the process and the innovative integration of machine learning and Petri nets. In recent years, more and more scholars have begun to pay attention to the medical field. At present, some researches apply Petri nets to the cellular level or introduce Petri nets into machine learning to improve the accuracy of machine learning. But our work is mainly to propose the novel research idea and framework for analysis of current multi-factor diseases evolutionary process and prove its feasibility. Every method has its own advantages. The advantages of this method lie in the fusion of CPN and machine learning in the field of multi-factor disease evolutionary process analysis. This new research idea can visualize the specific changes of different influencing factors, and analyze the data of different individuals, thus achieving the coexistence of personalization and visualization. At the same time, this model also has some limitations, which are mainly reflected in the fact that due to the particularity of depression itself, we could not obtain the real hormone data, so there are certain deviations in data results. Our major contribution is the novel idea of formal modeling and analysis of the evolution process of multi-factor diseases. The integration of CPN and machine learning will be further explored in future studies. With regard to the model optimization, we plan to use a more integrated system to combine the data analysis part of machine learning and the modeling part of CPN, so as to make the model building more convenient. At the same time, the model structure can be optimized to explore more links among internal factors of the disease.

5. Conclusion

This paper studies the interactions between various factors in diseases, and proposes a novel approach that integrates Petri nets and machine learning to model multi-factorial disorders. The key points of this method are those the model can demonstrate the disease evolutionary process, and also visualize the process of disease changes. It can better show the evolutionary process and development trend of diseases, and intuitively shows the results of multi-factor disease analysis. Absolutely, we built a MDP based on CPN, which allows us to express characteristic data as tokens. MDP can dynamically simulate in *CPNTools*, and output the probability of disease under different factors. Then, we apply our method in a multi-factorial analysis of depression, and proposed MDP model of depression to predict the effects of

serotonin, norepinephrine, and dopamine hormones. In addition, the MDP abstract model proposed in this paper can also be used in the analysis of other multi-factor diseases. It will better shows the evolutionary process and development trend of diseases, and intuitively shows the results of multi-factor disease analysis. At the same time, it can also provide better predictive assistance for clinical medical diagnosis. In future, we will continue to carry out in-depth research and optimize the multi-factor disease analysis model to make it more widely applicable and perform better. We will also explore further aspects of machine learning, such as deep learning, which may be suitable for reserch the evolutionary computation process of multi-factorial diseases.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was in part supported by the Natural Science Foundation of Shaanxi Province under Grants 2021JM-205, and the Fundamental Research Funds for the Central Universities.

References

[1] P. B. Bhattad and V. Jain. Artificial intelligence in modern medicine the evolving necessity of the present and role in transforming the future of medical care. *Cureus*, 12(5), 2020. doi:10.7759/cureus.8041.

[2] Y. K. Chan, Y. F. Chen, P. Tuan, W. Chang, and H. Ming-Yuan. Artificial intelligence in medical applications. *Journal of Healthcare Engineering*, 2018:1–2, 2018. doi:10.1155/2018/4827875.

[3] V. Kamra, P. Kumar, and M. Mohammadian. Formulation of an elegant diagnostic approach for an intelligent disease recommendation system. In *2019 9th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pages 278–281, 2019. doi:10.1109/CONFLUENCE.2019.8776952.

[4] K. Paranjape, M. Schinkel, and P. Nanayakkara. Short keynote paper: Mainstreaming personalized healthcare–transforming healthcare through new era of artificial intelligence. *IEEE Journal of Biomedical and Health Informatics*, 24(7):1860–1863, 2020. doi:10.1109/JBHI.2020.2970807.

[5] M. Nadji-Tehrani and A. Eslami. A brain-inspired framework for evolutionary artificial general intelligence. *IEEE Transactions on Neural Networks and Learning Systems*, 31(12):5257–5271, 2020. doi:10.1109/TNNLS.2020.2965567.

[6] C. C. Black, R. Richards, and J. M. Czum. Educational case: Aspiration pneumonia. *Academic Pathology*, 6:2374289519893082, 2019. doi:10.1109/TNNLS.2020.2965567.

[7] P. Chen, S. Yi, M. Ge, M. Wang, and L. Ao. Discovering key regulatory mechanisms from single-factor and multi-factor regulations in glioblastoma utilizing multi-dimensional data. *Molecular Biosystems*, 11(8):2345–2353, 2015. doi:10.1038/s41598-017-15890-9.

[8] D.A. Loewenstein, A. Acevedo, R. Ownby, and L. Schram. The nincds-adrda neuropsychological criteria for alzheimer's disease: a confirmatory factor analysis of single versus multi-factor models. *Archives of Clinical Neuropsychology*, 15(8):747, 2000. doi:10.1016/S0887-6177(00)80181-X.

[9] Y.M. Liu, X.Y. Zhang, W.W. Lin, N. Kehriman, W. Kuang, and X.M. Ling. Multi-factor combined biomarker screening strategy to rapidly diagnose alzheimer's disease and evaluate drug effect based on a rat model. *Journal of Pharmaceutical Analysis*, 12(4):627–636, 2022. doi:10.1016/j.jpha.2022.04.003.

[10] G. Ciriello and L. Magnani. The many faces of cancer evolution. *iScience*, 24(5):102403, 2021. doi:10.1016/j.isci.2021.102403.

[11] K. Jensen, L. M. Kristensen, and L. Wells. Coloured petri nets and cpn tools for modelling and validation of concurrent systems. *International Journal on Software Tools for Technology Transfer*, 9(3-4):213–254, 2007. doi:10.1007/s10009-007-0038-x.

[12] Z.H. Wu. *Petri Nets Introduction*. Machinery Industry Press, 2006.

[13] G. Yan and Y. Chen. Depression: fourth most common disorder worldwide. *Xinmin Weekly*, 30:4, 2014.

[14] X.R Wu, K.K Zhang, Y.L Xing, W. Zhou, Y.Q. Shao, G.Z Li, and Q. Rui. Dysregulated thyroid hormones correlate with anxiety and depression risk in patients with autoimmune disease. *Journal of Clinical Laboratory Analysis*, 35(1):e23573, 2021. doi:10.1002/jcla.23573.

[15] W.Y. Yu, M.H. Jia, X.W Fang, Y. Lu, and J.C Xu. Modeling and analysis of medical resource allocation based on timed colored petri net. *Future Generation Computer Systems*, 111:368–374, 2020. doi:10.1016/j.future.2020.05.010.

[16] S. Gupta, Z. Fatima, and S. Kumawat. Study of the bioenergetics to identify the novel pathways as a drug target against mycobacterium tuberculosis using petri net. *Biosystems*, 209(2):104509, 2021. doi:10.1016/j.biosystems.2021.104509.

[17] M. Mehraei and N. I. Akcay. Mood and emotional states prediction by time series methods. In *International Psychological Applications Conference and Trends*, 2016.

[18] M. Mehraei and N. I. Akcay. Pleasure, arousal, and dominance mood traits prediction using time series methods. *IAFOR Journal of Psychology and the Behavioral Sciences*, 3(1), 2017. doi:10.22492/ijpbs.3.1.01.

[19] I. Kawamitsu and M. Nakamura. *Colored Petri Net Modeling for Prediction Processes in Machine Learning*. Trends in Artificial Intelligence Theory and Applications. Artificial Intelligence Practices, 33rd International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2020, Kitakyushu, Japan, September 22-25, 2020, Proceedings, 2020.

[20] M. Mehraei. Identifying emotion regulation altering targets as depressive mood disorder treatments using fuzzy stochastic hybrid petri nets. *IAFOR Journal of Psychology and the Behavioral Sciences*, 4(1), 2018. doi:10.22492/IJPBS.4.1.04.

[21] M. Mehraei. Investigating the role of sleeping time in depressive mood disorders using simulation-based quantitative methods. *Modern Technology and Enginerring*, 2020.

[22] M. Mehraei. Mood states prediction by stochastic petri nets. In *International Psychological Applications Conference and Trends 2017*, 2017.

[23] M. K. Daoud and M. Otair. The role of artificial intelligence and the internet of things in the development of medical radiology (an experimental study on magnetic resonance imaging). In *2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI)*, 2020.

[24] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. *IEEE Reviews in Biomedical Engineering*, 14(99):4–15, 2020. doi:10.1109/RBME.2020.2987975.

[25] A. Panayides, A. Amini, N. Filipovic, A. Sharma, and C. Pattichis. Ai in medical imaging informatics: Current challenges and future directions. *IEEE Journal of Biomedical and Health Informatics*, 24(7):1837–1857, 2020. doi:10.1109/JBHI.2020.2991043.

[26] M. Nauman, N. Akhtar, A. Alhudhaif, and A. Alothaim. Guaranteeing correctness of machine learning based decision making at higher educational institutions. *IEEE Access*, 9:92864–92880, 2021. doi:10.1109/ACCESS.2021.3088901.

[27] H. Jiang and O. Nachum. Identifying and correcting label bias in machine learning. *23rd International Conference on Artificial Intelligence and Statistics, Online Meeting*, pages 702–712, 2020.

[28] Christoph M. Interpretable machine learning. *E-book*, 2022. `https://christophm.github.io/interpretable-ml-book/index.html`.

[29] W.Y. Yu, C.G. Yan, Z.J. Ding, C.J. Jiang, and M.C. Zhou. Modeling and validating e-commerce business process based on petri nets. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(3):327–341, 2014. doi:10.1109/TSMC.2013.2248358.

[30] W.Y. Yu, C.G. Yan, Z.J. Ding, C.J. Jiang, and M.C. Zhou. Modeling and verification of online shopping business processes by considering malicious behavior patterns. *IEEE Transactions on Automation Science and Engineering*, 13(2):647–662, 2016. doi:10.1109/TASE.2014.2362819.

[31] W.Y. Yu, C.G. Yan, Z.J. Ding, C.J. Jiang, and M.C. Zhou. Analyzing e-commerce business process nets via incidence matrix and reduction. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 48(1):130–141, 2018. doi:10.1109/TSMC.2016.2598287.

[32] Y.D Wang, W.Y Yu, P. Teng, G.J. Liu, and D.M. Xiang. A detection method for abnormal transactions in e-commerce based on extended data flow conformance checking. *Wireless Communications and Mobile Computing*, pages 1530–8669, 2022. doi:10.1155/2022/4434714.

[33] X.L. Chen, H. Peng, J. Wang, and F. Hao. Supervisory control of discrete event systems under asynchronous spiking neuron p systems. *Information Sciences*, 597:253–273, 2022. doi:10.1016/j.ins.2022.03.003.

[34] A Ja, A Kk, A Gv, A Im, B Kga, A Ks, and A Mj. Subclinical hypothyroidism and symptoms of depression: Evidence from the national health and nutrition examination surveys (nhanes). *Comprehensive Psychiatry*, 2021.

[35] X. Zhao, S.N.and Zhang and Y.L. Zhou. Comparison of thyroid function in different emotional states of drug-naive patients with bipolar disorder. *BMC ENDOCRINE DISORDERS*, 21(1):1–1, 2021. doi:10.1186/s12902-021-00869-5.

[36] Habib and Erensoy. The association between anxiety and depression with 25(oh)d and thyroid stimulating hormone levels. *Neurosciences (Riyadh, Saudi Arabia)*, 24(4):290–295, 2019. doi:10.17712/nsj.2019.4.20190028.

[37] Y. Zheng, T. Deng, and Y. Wang. Autism classification based on logistic regression model. In *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, 2021.

[38] S. Menard. Applied logistic regression analysis. thousand oaks, ca: Sage, 1995. 1995.

[39] L. Wells, H. M. Lassen, K. B. Lassen, M. Laursen, and K. Jensen. Cpn tools for editing, simulating, and analysing coloured petri nets - version 1.0 and beyond. 2003.