CLAIMS

1. A disease prediction method based on the long short-term memory (LSTM) model, comprising:

obtaining first medical data of a target object and second medical data of an associated object, wherein there is a blood relationship between the target object and the associated object; the first medical data comprises a medication history, a disease history, and a surgery history; the second medical data comprises a genetic disease treatment history;

inputting the first medical data and the second medical data into a first LSTM network in the trained LSTM model for operation, to obtain a hidden state vector sequence in the first LSTM network, wherein the LSTM model comprises the first LSTM network for encoding and a second LSTM network for decoding;

inputting the hidden state vector sequence into the second LSTM network for operation, to obtain a disease prediction result, wherein the disease prediction result comprises a predicted disease type and a corresponding incidence rate;

selecting a predicted disease with an incidence rate higher than a preset threshold from the disease prediction result, and recording the predicted disease as a designated disease, and obtaining, based on a preset disease association network, an associated disease directly connected to the designated disease, wherein network nodes of the association network are different types of diseases; and

outputting the disease prediction result and the associated disease;

wherein the step of inputting the first medical data and the second medical data into the first LSTM network in the trained LSTM model for operation, to obtain the hidden state vector sequence in the first LSTM network comprises:

dividing the first medical data into multiple data sequences based on preset time periods;

obtaining a designated impact factor of the genetic disease in the second medical data on other diseases based on a preset correspondence between the genetic disease and impact factors of other diseases; and

inputting the multiple data sequences and the designated impact factor into the first LSTM network in the trained LSTM model for operation, to obtain the hidden state vector sequence in the first LSTM network;

wherein the step of inputting the multiple data sequences and the designated impact factor into the first LSTM network in the trained LSTM model for operation, to obtain the hidden state vector sequence in the first LSTM network comprises:

obtaining the hidden state vector h_t in the first LSTM network according to the following formula: $h_t = LSTM_{enc}(x_t, h_{t-1})$, wherein t represents the t-th time period; h_t represents a hidden state vector corresponding to the t-th time period; h_{t-1} represents a hidden state vector corresponding to the (t–1)-th time period; X_t represents input data in the t-th time period; and $LSTM_{enc}$ refers to an encoding operation using the first LSTM network, wherein X_t comprises first medical data corresponding to the t-th time period and a designated impact factor corresponding to the t-th time period; and

constructing a hidden state vector sequence h_1, h_2, ..., h_n by using hidden state vectors corresponding to multiple preset time periods, wherein there are n time periods in total;

wherein the step of constructing a hidden state vector sequence h_1, h_2, ..., h_n by using hidden state vectors corresponding to multiple preset time periods, wherein there are n time periods in total, comprises:

obtaining the final hidden state vector c_i in the first LSTM network according to the following formulas: $c_i = \sum_{j=1}^{n} \alpha_{ij} h_j$, $\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{n} \exp(e_{ik})}$, and $e_{ij} = score(s_i, h_j)$, wherein a_{ij} represents a weighting parameter, and there are n time periods in total; s_i represents the i-th hidden state vector in the second LSTM network; $score(s_i, h_j)$ represents a score calculated based on s_i and h_j using a preset score function; and

constructing a hidden state vector sequence c_1, c_2, ..., c_n by using final hidden state vectors corresponding to multiple preset time periods.

2. The disease prediction method based on the LSTM model according to claim 1, wherein the step of inputting the hidden state vector sequence into the second LSTM network for operation, to obtain a disease prediction result, wherein the disease prediction result comprises a predicted disease type and a corresponding incidence rate, comprises:

inputting the hidden state vector sequence into the second LSTM network for operation, to obtain a high-dimensional vector sequence that is output by the second LSTM network; and

interpreting the high-dimensional vector sequence based on a preset correspondence between a component vector and a meaning of the prediction result, so as to obtain disease prediction results

in different time periods in the future, wherein the disease prediction result comprises the predicted disease type and the corresponding incidence rate.

3. The disease prediction method based on the LSTM model according to claim 1, wherein after the step of inputting the hidden state vector sequence into the second LSTM network for operation, to obtain a disease prediction result, wherein the disease prediction result comprises a predicted disease type and a corresponding incidence rate, the method comprises:

receiving multiple input improvement factor groups, and inputting the improvement factor groups, the first medical data, and the second medical data into the trained LSTM model for calculation, wherein the improvement factor groups comprise carrying out of medication or surgery at designated time points;

obtaining multiple groups of improved disease prediction results respectively corresponding to the multiple improvement factor groups output by the LSTM model, wherein the improved disease prediction results comprise predicted disease types and corresponding incidence rates; and

selecting a final improved disease prediction result from the multiple groups of improved disease prediction results based on a preset selection rule, and generating a recommended treatment plan, wherein the recommended treatment plan is accompanied by the improvement factor group corresponding to the final improved disease prediction result.

4. The disease prediction method based on the LSTM model according to claim 1, wherein the disease association network is a knowledge graph network; before the step of obtaining, based on the preset disease association network, the associated disease directly connected to the designated disease, wherein the network nodes of the association network are different types of diseases, the method comprises:

using a preset knowledge graph construction tool to identify initial entities from designated information collected in advance, wherein the designated information records at least the designated disease, and the initial entities comprise at least the designated disease;

deduplicating the initial entities to obtain final entities; and

extracting a relationship between the final entities from the designated information to form a triplet, and generating the knowledge graph network based on the triplet.

5. A disease prediction apparatus based on the LSTM model, comprising:

a medical data acquisition unit, configured to obtain first medical data of a target object and

second medical data of an associated object, wherein there is a blood relationship between the target object and the associated object; the first medical data comprises a medication history, a disease history, and a surgery history; the second medical data comprises a genetic disease treatment history;

a hidden state vector sequence acquisition unit, configured to input the first medical data and the second medical data into a first LSTM network in the trained LSTM model for operation, to obtain a hidden state vector sequence in the first LSTM network, wherein the LSTM model comprises the first LSTM network for encoding and a second LSTM network for decoding;

a disease prediction result acquisition unit, configured to input the hidden state vector sequence into the second LSTM network for operation, to obtain a disease prediction result, wherein the disease prediction result comprises a predicted disease type and a corresponding incidence rate;

an associated disease acquisition unit, configured to select a predicted disease with an incidence rate higher than a preset threshold from the disease prediction result, and record the predicted disease as a designated disease, and obtain, based on a preset disease association network, an associated disease directly connected to the designated disease, wherein network nodes of the association network are different types of diseases; and

an output unit, configured to output the disease prediction result and the associated disease;

wherein the hidden state vector sequence acquisition unit comprises:

a multiple data sequence acquisition subunit, configured to divide the first medical data into multiple data sequences based on preset time periods;

a designated impact factor acquisition subunit, configured to obtain a designated impact factor of the genetic disease in the second medical data on other diseases based on a preset correspondence between the genetic disease and impact factors of other diseases; and

a hidden state vector sequence acquisition subunit, configured to input the multiple data sequences and the designated impact factor into the first LSTM network in the trained LSTM model for operation, to obtain the hidden state vector sequence in the first LSTM network;

wherein the hidden state vector sequence acquisition subunit comprises:

a hidden state vector calculation module, configured to obtain the hidden state vector h_t in the first LSTM network according to the following formula: $h_t = LSTM_{enc}(x_t, h_{t-1})$, wherein t

represents the t-th time period; h_t represents a hidden state vector corresponding to the t-th time period; h_{t-1} represents a hidden state vector corresponding to the (t–1)-th time period; X_t represents input data in the t-th time period; and $LSTM_{enc}$ refers to an encoding operation using the first LSTM network, wherein X_t comprises first medical data corresponding to the t-th time period and a designated impact factor corresponding to the t-th time period; and

a hidden state vector sequence acquisition module, configured to construct a hidden state vector sequence $h_1, h_2, ..., h_n$ by using hidden state vectors corresponding to multiple preset time periods, wherein there are n time periods in total;

wherein the hidden state vector sequence acquisition module comprises:

a final hidden state vector acquisition submodule, configured to obtain the final hidden state vector c_i in the first LSTM network according to the following formulas: $c_i = \sum_{j=1}^{n} \alpha_{ij} h_j$, $\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{n} \exp(e_{ik})}$, and $e_{ij} = score(s_i, h_j)$, wherein a_{ij} represents a weighting parameter, and there are n time periods in total; s_i represents the i-th hidden state vector in the second LSTM network; $score(s_i, h_j)$ represents a score calculated based on s_i and h_j using a preset score function; and

a hidden state vector sequence acquisition submodule, configured to construct a hidden state vector sequence $c_1, c_2, ..., c_n$ by using final hidden state vectors corresponding to multiple preset time periods.

6. The disease prediction apparatus based on the LSTM model according to claim 5, wherein the disease prediction result acquisition unit comprises:

a high-dimensional vector sequence acquisition subunit, configured to input the hidden state vector sequence into the second LSTM network for operation, to obtain a high-dimensional vector sequence that is output by the second LSTM network; and

a disease prediction result acquisition subunit, configured to interpret the high-dimensional vector sequence based on a preset correspondence between a component vector and a meaning of the prediction result, so as to obtain disease prediction results in different time periods in the future, wherein the disease prediction result comprises the predicted disease type and the corresponding incidence rate.

7. The disease prediction apparatus based on the LSTM model according to claim 5, wherein the apparatus comprises:

an improvement factor group receiving unit, configured to receive multiple input

improvement factor groups, and input the improvement factor groups, the first medical data, and the second medical data into the trained LSTM model for calculation, wherein the improvement factor groups comprise carrying out of medication or surgery at designated time points;

an improved disease prediction result acquisition unit, configured to obtain multiple groups of improved disease prediction results respectively corresponding to the multiple improvement factor groups output by the LSTM model, wherein the improved disease prediction results comprise predicted disease types and corresponding incidence rates; and

a recommended treatment plan generation unit, configured to select a final improved disease prediction result from the multiple groups of improved disease prediction results based on a preset selection rule, and generate a recommended treatment plan, wherein the recommended treatment plan is accompanied by the improvement factor group corresponding to the final improved disease prediction result.

8. The disease prediction apparatus based on the LSTM model according to claim 5, wherein the disease association network is a knowledge graph network, and the apparatus comprises:

an initial entity identification unit, configured to use a preset knowledge graph construction tool to identify initial entities from designated information collected in advance, wherein the designated information records at least the designated disease, and the initial entities comprise at least the designated disease;

a final entity acquisition unit, configured to deduplicate the initial entities to obtain final entities; and

a knowledge graph network generation unit, configured to extract a relationship between the final entities from the designated information to form a triplet, and generate the knowledge graph network based on the triplet.

9. A computer device, comprising a memory and a processor, wherein the memory stores a computer program, the processor executes the computer program to perform a disease prediction method based on the LSTM model, and the disease prediction method based on the LSTM model comprises:

obtaining first medical data of a target object and second medical data of an associated object, wherein there is a blood relationship between the target object and the associated object; the first medical data comprises a medication history, a disease history, and a surgery history; the second

medical data comprises a genetic disease treatment history;

inputting the first medical data and the second medical data into a first LSTM network in the trained LSTM model for operation, to obtain a hidden state vector sequence in the first LSTM network, wherein the LSTM model comprises the first LSTM network for encoding and a second LSTM network for decoding;

inputting the hidden state vector sequence into the second LSTM network for operation, to obtain a disease prediction result, wherein the disease prediction result comprises a predicted disease type and a corresponding incidence rate;

selecting a predicted disease with an incidence rate higher than a preset threshold from the disease prediction result, and recording the predicted disease as a designated disease, and obtaining, based on a preset disease association network, an associated disease directly connected to the designated disease, wherein network nodes of the association network are different types of diseases; and

outputting the disease prediction result and the associated disease;

wherein the step of inputting the first medical data and the second medical data into the first LSTM network in the trained LSTM model for operation, to obtain the hidden state vector sequence in the first LSTM network comprises:

dividing the first medical data into multiple data sequences based on preset time periods;

obtaining a designated impact factor of the genetic disease in the second medical data on other diseases based on a preset correspondence between the genetic disease and impact factors of other diseases; and

inputting the multiple data sequences and the designated impact factor into the first LSTM network in the trained LSTM model for operation, to obtain the hidden state vector sequence in the first LSTM network;

wherein the step of inputting the multiple data sequences and the designated impact factor into the first LSTM network in the trained LSTM model for operation, to obtain the hidden state vector sequence in the first LSTM network comprises:

obtaining the hidden state vector h_t in the first LSTM network according to the following formula: $h_t = LSTM_{enc}(x_t, h_{t-1})$, wherein t represents the t-th time period; h_t represents a hidden state vector corresponding to the t-th time period; h_{t-1} represents a hidden state vector

corresponding to the (t–1)-th time period; X_t represents input data in the t-th time period; and $LSTM_{enc}$ refers to an encoding operation using the first LSTM network, wherein X_t comprises first medical data corresponding to the t-th time period and a designated impact factor corresponding to the t-th time period; and

constructing a hidden state vector sequence h_1, h_2, ..., h_n by using hidden state vectors corresponding to multiple preset time periods, wherein there are n time periods in total;

wherein the step of constructing a hidden state vector sequence h_1, h_2, ..., h_n by using hidden state vectors corresponding to multiple preset time periods, wherein there are n time periods in total, comprises:

obtaining the final hidden state vector c_i in the first LSTM network according to the following formulas: $c_i = \sum_{j=1}^{n} \alpha_{ij} h_j$, $\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{n} \exp(e_{ik})}$, and $e_{ij} = score(s_i, h_j)$, wherein a_{ij} represents a weighting parameter, and there are n time periods in total; s_i represents the i-th hidden state vector in the second LSTM network; $score(s_i, h_j)$ represents a score calculated based on s_i and h_j using a preset score function; and

constructing a hidden state vector sequence c_1, c_2, ..., c_n by using final hidden state vectors corresponding to multiple preset time periods.

10. A computer readable storage medium, wherein the computer readable storage medium stores a computer program, the computer program is executed by a processor to perform a disease prediction method based on the LSTM model, and the disease prediction method based on the LSTM model comprises:

obtaining first medical data of a target object and second medical data of an associated object, wherein there is a blood relationship between the target object and the associated object; the first medical data comprises a medication history, a disease history, and a surgery history; the second medical data comprises a genetic disease treatment history;

inputting the first medical data and the second medical data into a first LSTM network in the trained LSTM model for operation, to obtain a hidden state vector sequence in the first LSTM network, wherein the LSTM model comprises the first LSTM network for encoding and a second LSTM network for decoding;

inputting the hidden state vector sequence into the second LSTM network for operation, to obtain a disease prediction result, wherein the disease prediction result comprises a predicted

disease type and a corresponding incidence rate;

selecting a predicted disease with an incidence rate higher than a preset threshold from the disease prediction result, and recording the predicted disease as a designated disease, and obtaining, based on a preset disease association network, an associated disease directly connected to the designated disease, wherein network nodes of the association network are different types of diseases; and

outputting the disease prediction result and the associated disease;

wherein the step of inputting the first medical data and the second medical data into the first LSTM network in the trained LSTM model for operation, to obtain the hidden state vector sequence in the first LSTM network comprises:

dividing the first medical data into multiple data sequences based on preset time periods;

obtaining a designated impact factor of the genetic disease in the second medical data on other diseases based on a preset correspondence between the genetic disease and impact factors of other diseases; and

inputting the multiple data sequences and the designated impact factor into the first LSTM network in the trained LSTM model for operation, to obtain the hidden state vector sequence in the first LSTM network;

wherein the step of inputting the multiple data sequences and the designated impact factor into the first LSTM network in the trained LSTM model for operation, to obtain the hidden state vector sequence in the first LSTM network comprises:

obtaining the hidden state vector h_t in the first LSTM network according to the following formula: $h_t = LSTM_{enc}(x_t, h_{t-1})$, wherein t represents the t-th time period; h_t represents a hidden state vector corresponding to the t-th time period; h_{t-1} represents a hidden state vector corresponding to the (t–1)-th time period; X_t represents input data in the t-th time period; and $LSTM_{enc}$ refers to an encoding operation using the first LSTM network, wherein X_t comprises first medical data corresponding to the t-th time period and a designated impact factor corresponding to the t-th time period; and

constructing a hidden state vector sequence h_1, h_2, ..., h_n by using hidden state vectors corresponding to multiple preset time periods, wherein there are n time periods in total;

wherein the step of constructing a hidden state vector sequence h_1, h_2, ..., h_n by using hidden

state vectors corresponding to multiple preset time periods, wherein there are n time periods in total, comprises:

obtaining the final hidden state vector c_i in the first LSTM network according to the following formulas: $c_i = \sum_{j=1}^{n} \alpha_{ij} h_j$, $\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{n} \exp(e_{ik})}$, and $e_{ij} = score(s_i, h_j)$, wherein a_{ij} represents a weighting parameter, and there are n time periods in total; s_i represents the i-th hidden state vector in the second LSTM network; $score(s_i, h_j)$ represents a score calculated based on s_i and h_j using a preset score function; and

constructing a hidden state vector sequence c_1, c_2, ..., c_n by using final hidden state vectors corresponding to multiple preset time periods.