Claims

1. A computerized method for predicting progression of chronic kidney disease for a patient using a machine learning model, the method comprising:

obtaining patient input data associated with the patient, the patient input data comprising:

an age of the patient;

a gender of the patient; and

a series of estimated glomerular filtration rates (eGFRs) of the patient, the series of eGFRs comprising a baseline eGFR and a plurality of historical eGFRs preceding the baseline eGFR;

obtaining a current patient input dataset associated with the patient, the current patient input dataset comprising:

the age of the patient;

the gender of the patient; and

the baseline eGFR from the series of eGFRs;

generating a patient input vector representation using the current patient input dataset and an input linear layer of the machine learning model;

obtaining a historical eGFR input dataset associated with the patient, the historical eGFR input dataset comprising the historical eGFRs from the series of eGFRs;

generating an eGFR vector representation using the historical eGFR input dataset and a first input machine learning module of the machine learning model;

concatenating the patient input and eGFR vector representations into an output vector representation; and

generating a prediction output using the output vector representation and an output machine learning module of the machine learning model, the prediction output comprising a risk assessment of the patient's progression of chronic kidney disease.

2. The method according to claim 1, wherein the first input machine learning module and the output machine learning module comprise neural networks.

3. The method according to claim 1 or 2, wherein the first input machine learning module comprises a gated recurrent unit neural network or a long short-term memory (LSTM) network, and the output machine learning module comprises a multilayer perceptron.

4. The method according to any one of claims 1 to 3, wherein the patient input data and the current patient input dataset comprise any one or more of: blood pressure, indication of whether the patient has diabetes and/or hypertension, blood urea nitrogen, serum sodium, fasting blood glucose, total cholesterol, urine creatinine, urine protein, urine protein creatinine ratio (PCR), urine albumin-creatinine ratio (uACR), serum calcium, serum phosphate, serum creatinine, serum chloride, serum potassium, serum bicarbonate, serum albumin, serum haemoglobin, C-reactive protein, total lipids, low-density lipoprotein (LDL), triglycerides, serum urea, haemoglobin A1C, urine phase contrast, serum uric acid, 24-hour urine volume, urea, serum iron, ferritin, transferrin saturation, haematocrit, intact parathyroid hormone (iPTH), aspartate transferase (AST), alanine transaminase (ALT), and alkaline phosphatase (ALP).

5. The method according to any one of claims 1 to 4,

wherein the patient input data comprises a series of measurements of a biochemical variable of the patient, the series of measurements of the biochemical variable comprising a baseline measurement of the biochemical variable and a plurality of historical measurements of the biochemical variable preceding the baseline measurement;

the method comprising:

obtaining a second historical input dataset associated with the patient, the second historical input dataset comprising the historical measurements of the biochemical variable,

wherein the current patient input dataset comprises the baseline measurement of the biochemical variable, the biochemical variable comprising any one of: blood urea nitrogen, serum sodium, fasting blood glucose, total cholesterol, urine creatinine, urine protein, urine protein creatinine ratio (PCR), urine albumin-creatinine ratio (uACR), serum calcium, serum phosphate, serum

creatinine, serum chloride, serum potassium, serum bicarbonate, serum albumin, serum haemoglobin, C-reactive protein, total lipids, low-density lipoprotein (LDL), triglycerides, serum urea, haemoglobin A1C, urine phase contrast, serum uric acid, 24-hour urine volume, urea, serum iron, ferritin, transferrin saturation, haematocrit, intact parathyroid hormone (iPTH), aspartate transferase (AST), alanine transaminase (ALT), and alkaline phosphatase (ALP); and

the method further comprising:

generating a second vector representation using the second historical input dataset and a second input machine learning module of the machine learning model; and

concatenating the patient input, eGFR, and second vector representations into the output vector representation.

6. The method according to any one of claims 1 to 5, wherein the progression of chronic kidney disease is associated with:

an eGFR gradient of at least 5 mL/min/1.73m^2 per year;

a decrease in eGFR of at least 40% over at least 3 months; or

sustained eGFR less than or equal to 15 mL/min/1.73m^2, and

wherein the machine learning model is trained using weights and/or biases obtained from another pre-trained machine learning model.

7. The method according to any one of claims 1 to 5, wherein the progression of chronic kidney disease is such that the patient is required to undergo maintenance dialysis or kidney transplant, and wherein the method further comprises determining weights and/or biases from the machine learning model to expedite training of another machine learning model.

8. The method according to any one of claims 1 to 7, further comprising generating a prediction report for visualization of the risk assessment, wherein the prediction report optionally comprises a risk classification for the patient where the patient is classified into one among a plurality of the risk classifications based on predefined risk thresholds.

9. The method according to claim 8, wherein the prediction report comprises explanations generated using an explainable artificial intelligence, and wherein the explainable artificial intelligence optionally comprises a Shapley Additive Explanations (SHAP) method and/or a Local Interpretable Model-Agnostic Explanations (LIME) method.

10. The method according to claim 8 or 9, wherein the prediction report comprises rankings of data variables in the current patient input dataset, the data variables ranked based on contribution to the prediction of the progression of chronic kidney disease, and wherein the prediction report optionally comprises a hypothetical risk assessment of the patient's progression of chronic kidney disease when one or more of the ranked data variables are modified.

11. A computer system comprising:

a machine learning model;

a processor configured for executing computer-readable instructions to predict progression of chronic kidney disease for a patient using the machine learning model; and

a patient database comprising patient input data associated with the patient, the patient input data comprising:

an age of the patient;

a gender of the patient; and

a series of estimated glomerular filtration rates (eGFRs) of the patient, the series of eGFRs comprising a baseline eGFR and a plurality of historical eGFRs preceding the baseline eGFR,

wherein the machine learning model comprises:

a patient input module configured for obtaining a current patient input dataset associated with the patient, the current patient input dataset comprising:

the age of the patient;

the gender of the patient; and

the baseline eGFR from the series of eGFRs;

an input linear layer configured for generating a patient input vector representation using the current patient input dataset;

a historical eGFR input module configured for obtaining a historical eGFR input dataset associated with the patient, the historical eGFR input dataset comprising the historical eGFRs from the series of eGFRs;

a first input machine learning module configured for generating an eGFR vector representation using the historical eGFR input dataset;

a vector concatenation module configured for concatenating the patient input and eGFR vector representations into an output vector representation; and

an output machine learning module configured for generating a prediction output using the output vector representation, the prediction output comprising a risk assessment of the patient's progression of chronic kidney disease.

12. A computerized method for training a machine learning model for predicting progression of chronic kidney disease for patients, the method comprising:

obtaining patient training data associated with a population of patients, the patient training data comprising, for each patient in the population, :

an age of the respective patient;

a gender of the respective patient; and

a series of estimated glomerular filtration rates (eGFRs) of the respective patient, the series of eGFRs comprising a baseline eGFR and a plurality of historical eGFRs preceding the baseline eGFR;

obtaining current patient training datasets associated with the population of patients, each current patient training dataset associated with a respective patient in the population and comprising:

the age of the respective patient;

the gender of the respective patient; and

the baseline eGFR from the respective series of eGFRs;

generating, for each current patient training dataset, a patient input vector representation using the respective current patient training dataset and an input linear layer;

obtaining historical eGFR training datasets associated with the population, each historical eGFR training dataset associated with a respective patient in the population and comprising the historical eGFRs from the respective series of eGFRs;

generating, for each historical eGFR training dataset, an eGFR vector representation using the respective historical eGFR training dataset and a first input machine learning module of the machine learning model;

concatenating, for each patient, the respective patient input and eGFR vector representations into a respective output vector representation; and

inputting the respective output vector representation and medical data of each patient into an output machine learning module of the machine learning model to thereby train the machine learning model, the medical data comprising details of the respective patient's progression of chronic kidney disease,

wherein the trained machine learning model is configured for generating a prediction output for a new patient, the prediction output comprising a risk assessment of the new patient's progression of chronic kidney disease.

13. The method according to claim 12, wherein the first input machine learning module and the output machine learning module comprise neural networks.

14. The method according to claim 12 or 13, wherein the first input machine learning module comprises a gated recurrent unit neural network or a long short-term memory (LSTM) network, and wherein the output machine learning module comprises a feedforward neural network or a multilayer perceptron.

15. The method according to any one of claims 12 to 14, wherein the series of eGFRs and the medical data further comprise a first eGFR succeeding the baseline eGFR.

16. The method according to any one of claims 12 to 15, wherein the patient training data and each current patient training dataset comprise any one or more of: blood pressure, indication of whether the patient has diabetes and/or hypertension, blood urea nitrogen, serum sodium, fasting blood glucose, total cholesterol, urine creatinine, urine protein, urine protein creatinine ratio (PCR), urine albumin-creatinine ratio (uACR), serum calcium, serum phosphate, serum creatinine, serum chloride, serum potassium, serum bicarbonate, serum albumin, serum haemoglobin, C-reactive protein, total lipids, low-density lipoprotein (LDL), triglycerides, serum urea, haemoglobin A1C, urine phase contrast, serum uric acid, 24-hour urine volume, urea, serum iron, ferritin, transferrin saturation, haematocrit, intact parathyroid hormone (iPTH), aspartate transferase (AST), alanine transaminase (ALT), and alkaline phosphatase (ALP).

17. The method according to any one of claims 12 to 16,

wherein the patient training data comprises a series of measurements of a biochemical variable of each patient, the series of measurements of the biochemical variable comprising a baseline measurement of the biochemical variable and a plurality of historical measurements of the biochemical variable preceding the baseline measurement;

the method comprising:

obtaining second historical training datasets associated with the population of patients, each second historical training dataset associated with a respective patient in the population and comprising the respective historical measurements of the biochemical variable,

wherein the respective current patient training dataset comprises the baseline measurement of the biochemical variable, the biochemical variable comprising any one of: blood urea nitrogen, serum sodium, fasting blood glucose, total cholesterol, urine creatinine, urine protein, urine protein creatinine ratio (PCR), urine albumin-creatinine ratio (uACR), serum calcium, serum phosphate, serum creatinine, serum chloride, serum potassium, serum bicarbonate, serum albumin, serum haemoglobin, C-reactive protein, total lipids, low-density lipoprotein (LDL), triglycerides, serum urea, haemoglobin A1C, urine phase contrast, serum uric acid, 24-hour urine volume, urea, serum iron,

ferritin, transferrin saturation, haematocrit, intact parathyroid hormone (iPTH), aspartate transferase (AST), alanine transaminase (ALT), and alkaline phosphatase (ALP); and

the method further comprising:

generating, for each second historical training dataset, a second vector representation using the respective second historical input dataset and a second input machine learning module of the machine learning model; and

concatenating the respective patient input, eGFR, and second vector representations into the respective output vector representation.

18. The method according to any one of claims 12 to 17, wherein the progression of chronic kidney disease is associated with:

an eGFR gradient of at least 5 mL/min/1.73m^2 per year;

a decrease in eGFR of at least 40% over at least 3 months; or

sustained eGFR less than or equal to 15 mL/min/1.73m^2, and

wherein the method further comprises training the machine learning model using weights and/or biases obtained from another machine learning model that has been trained using the same patient training datasets and historical eGFR training datasets.

19. The method according to any one of claims 12 to 17, wherein the progression of chronic kidney disease is such that the respective patient is required to undergo maintenance dialysis or kidney transplant, and wherein the method further comprises determining weights and/or biases from the machine learning model to expedite training of another machine learning model with the same current patient training datasets and historical eGFR training datasets.

20. A computer system comprising:

a machine learning model;

a processor configured for executing computer-readable instructions to train the machine learning model for predicting progression of chronic kidney disease for patients; and

a patient database comprising patient input data associated with a population of patients, the patient input data comprising, for each patient in the population, :

an age of the respective patient;

a gender of the respective patient; and

a series of estimated glomerular filtration rates (eGFRs) of the respective patient, the series of eGFRs comprising a baseline eGFR and a plurality of historical eGFRs preceding the baseline eGFR,

wherein the machine learning model comprises:

a patient input module configured for obtaining current patient training datasets associated with the population of patients, each current patient training dataset associated with a respective patient in the population and comprising:

the age of the respective patient;

the gender of the respective patient; and

the baseline eGFR from the respective series of eGFRs;

an input linear layer configured for generating, for each current patient training dataset, a patient input vector representation using the respective current patient training dataset and an input linear layer;

a historical eGFR input module configured for obtaining historical eGFR training datasets associated with the population, each historical eGFR training dataset associated with a respective patient in the population and comprising the historical eGFRs from the respective series of eGFRs;

a first input machine learning module configured for generating, for each historical eGFR training dataset, an eGFR vector representation using the respective historical eGFR training dataset;

a vector concatenation module configured for concatenating, for each patient, the respective patient input and eGFR vector representations into a respective output vector representation; and

an output machine learning module configured for receiving the respective output vector representation and medical data of each patient

to thereby train the machine learning model, the medical data comprising details of the respective patient's progression of chronic kidney disease,

wherein the trained machine learning model is configured for generating a prediction output for a new patient, the prediction output comprising a risk assessment of the new patient's progression of chronic kidney disease.