



(19) **United States**

(12) **Patent Application Publication**  
**BRUE et al.**

(10) **Pub. No.: US 2021/0313068 A1**

(43) **Pub. Date: Oct. 7, 2021**

(54) **RECOMMENDING TREATMENTS TO MITIGATE MEDICAL CONDITIONS AND PROMOTE SURVIVAL OF LIVING ORGANISMS USING MACHINE LEARNING MODELS**

*G16H 50/20* (2006.01)  
*G16H 50/70* (2006.01)  
*G16H 20/10* (2006.01)  
*G16H 20/40* (2006.01)  
*G06N 5/04* (2006.01)  
*G06N 7/00* (2006.01)

(71) Applicant: **General Genomics, Inc.**, Midland, TX (US)

(52) **U.S. Cl.**  
CPC ..... *G16H 50/30* (2018.01); *G06N 20/00* (2019.01); *G16H 50/20* (2018.01); *G06N 7/005* (2013.01); *G16H 20/10* (2018.01); *G16H 20/40* (2018.01); *G06N 5/04* (2013.01); *G16H 50/70* (2018.01)

(72) Inventors: **Daniel Alan BRUE**, Edmond, OK (US); **Warren Dennis GIECK**, Calgary (CA); **Aronjol David ROSENTHAL**, Midland, TX (US)

(21) Appl. No.: **17/207,440**

(57) **ABSTRACT**

(22) Filed: **Mar. 19, 2021**

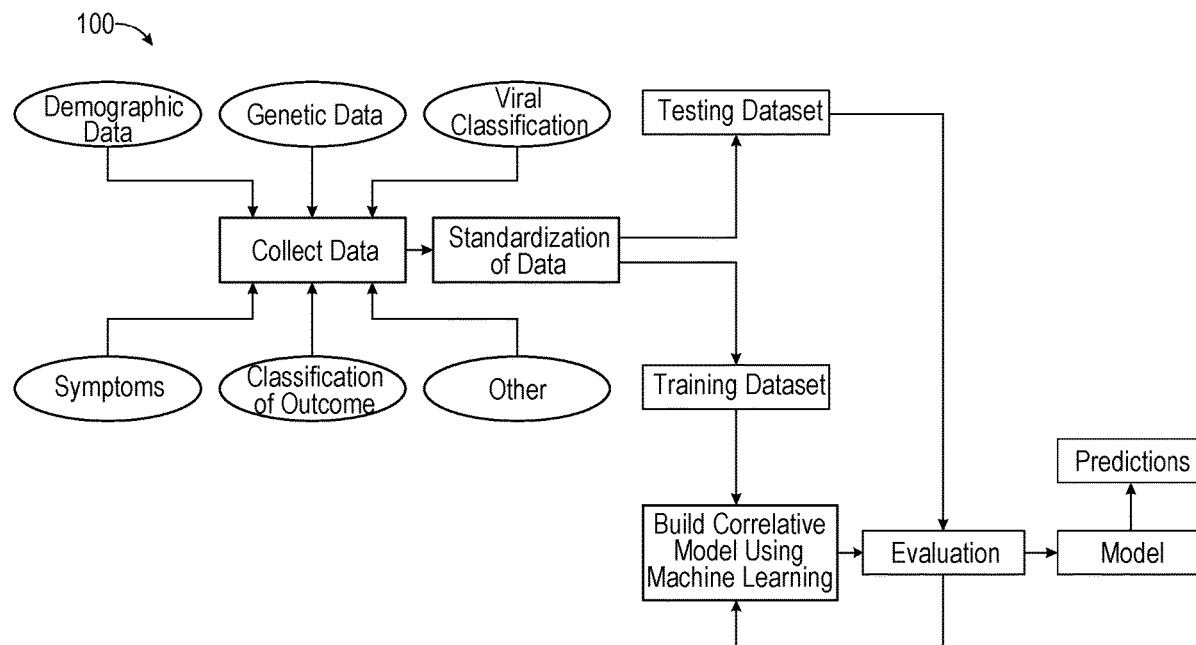
**Related U.S. Application Data**

Embodiments of the present disclosure generally relate to methods for analyzing survivability of illnesses, such as COVID-19. More particularly, embodiments of the present disclosure relate to methods for identifying correlations and influencing factors between genetic markers, lifestyle, and other available data that lead to predictions of the effectiveness of medical treatments, predicting results of mass exposure to an illness based on a population's genomes and other available data, and providing indicators and methods of visualization for survivability of a viral infection or cancer in any living organism.

(60) Provisional application No. 63/005,916, filed on Apr. 6, 2020.

**Publication Classification**

(51) **Int. Cl.**  
*G16H 50/30* (2006.01)  
*G06N 20/00* (2006.01)



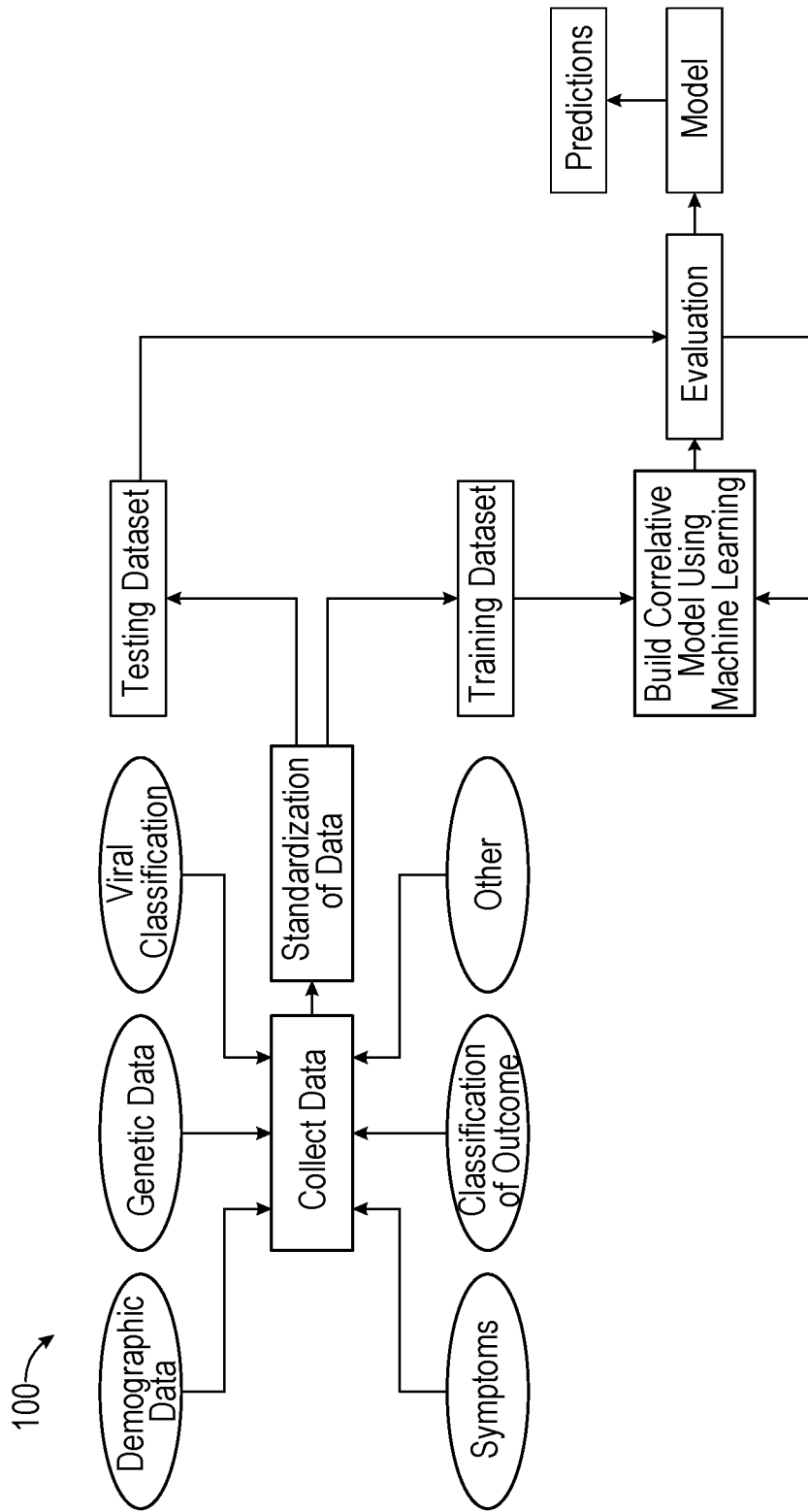


FIG. 1

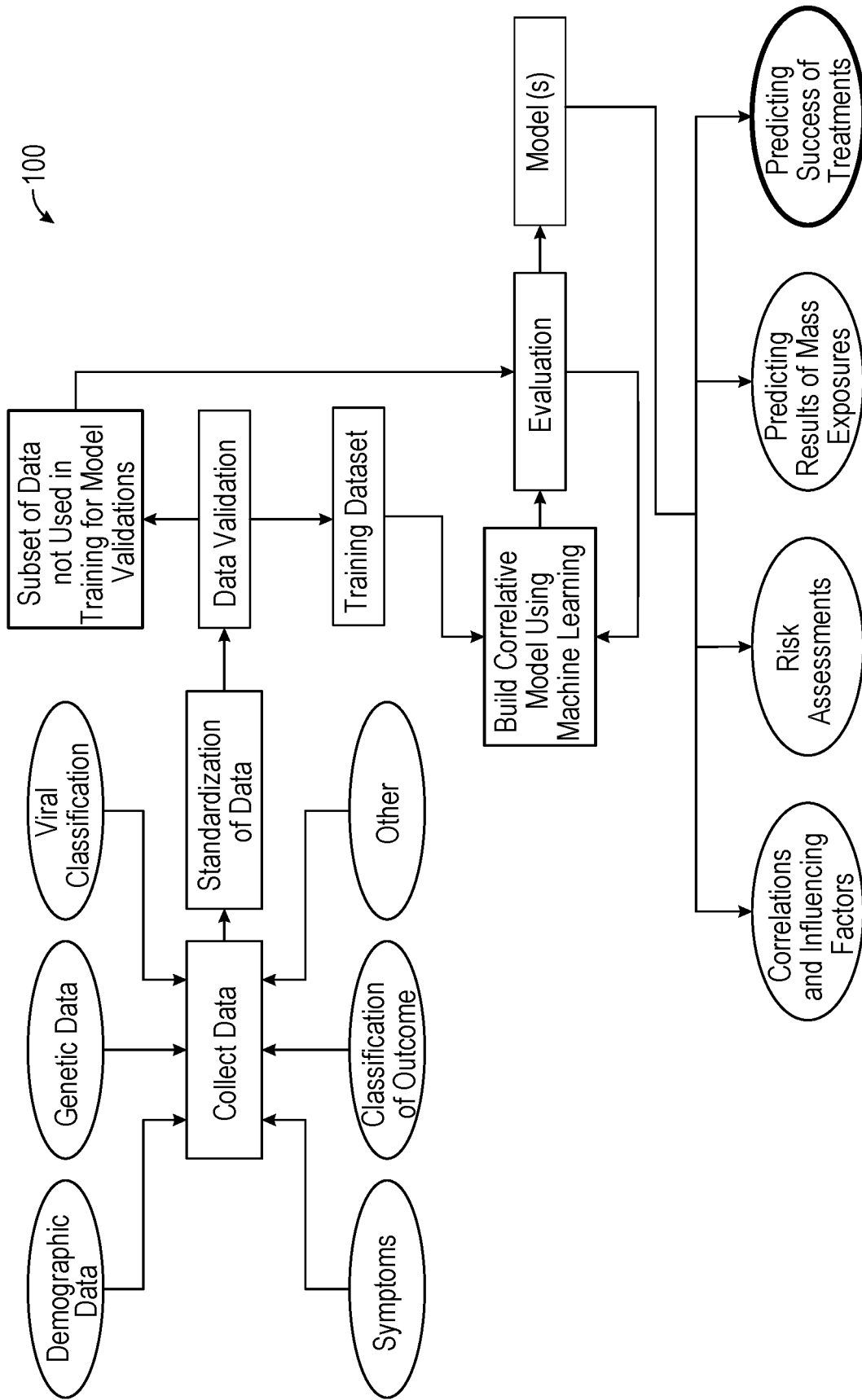


FIG. 2

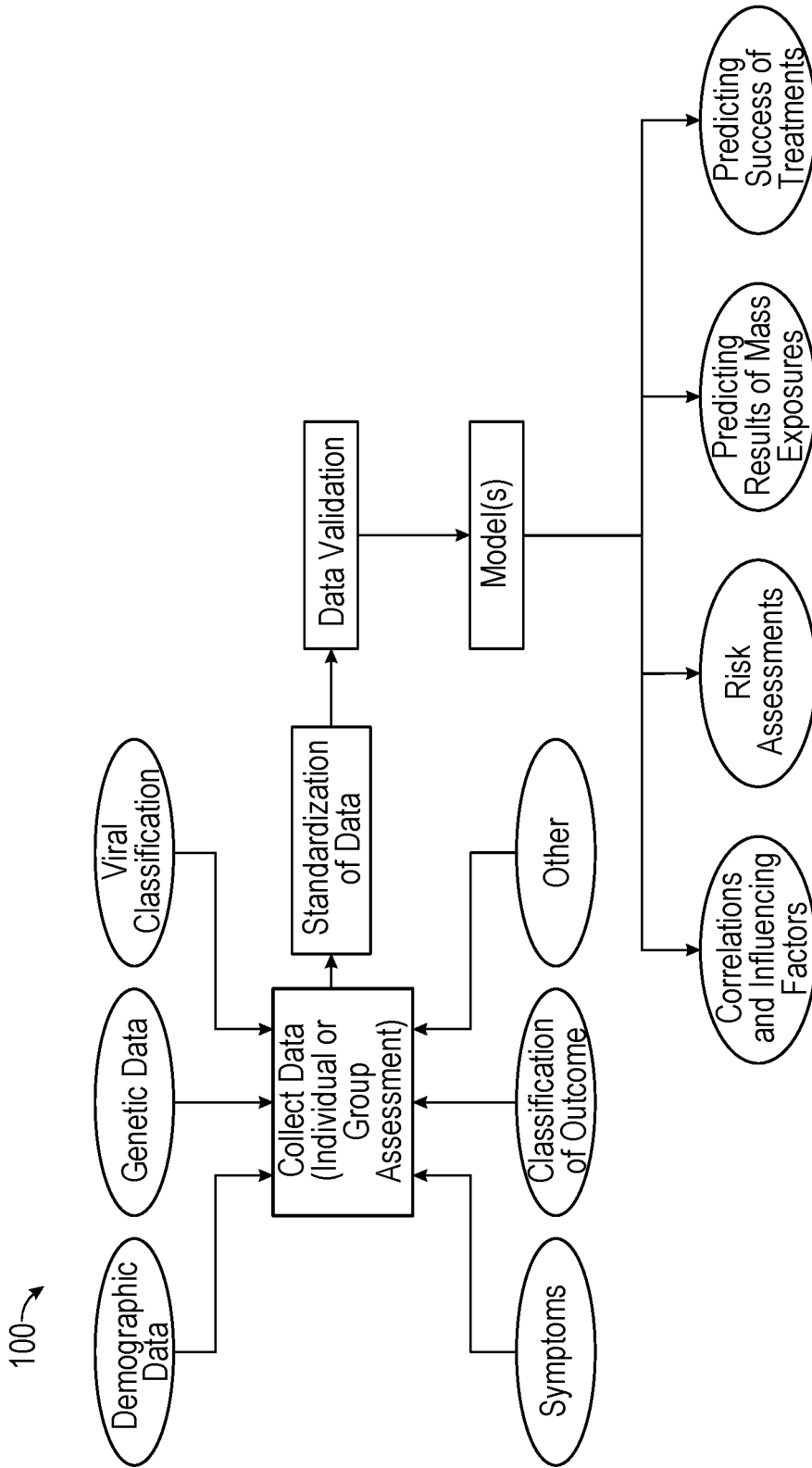


FIG. 3

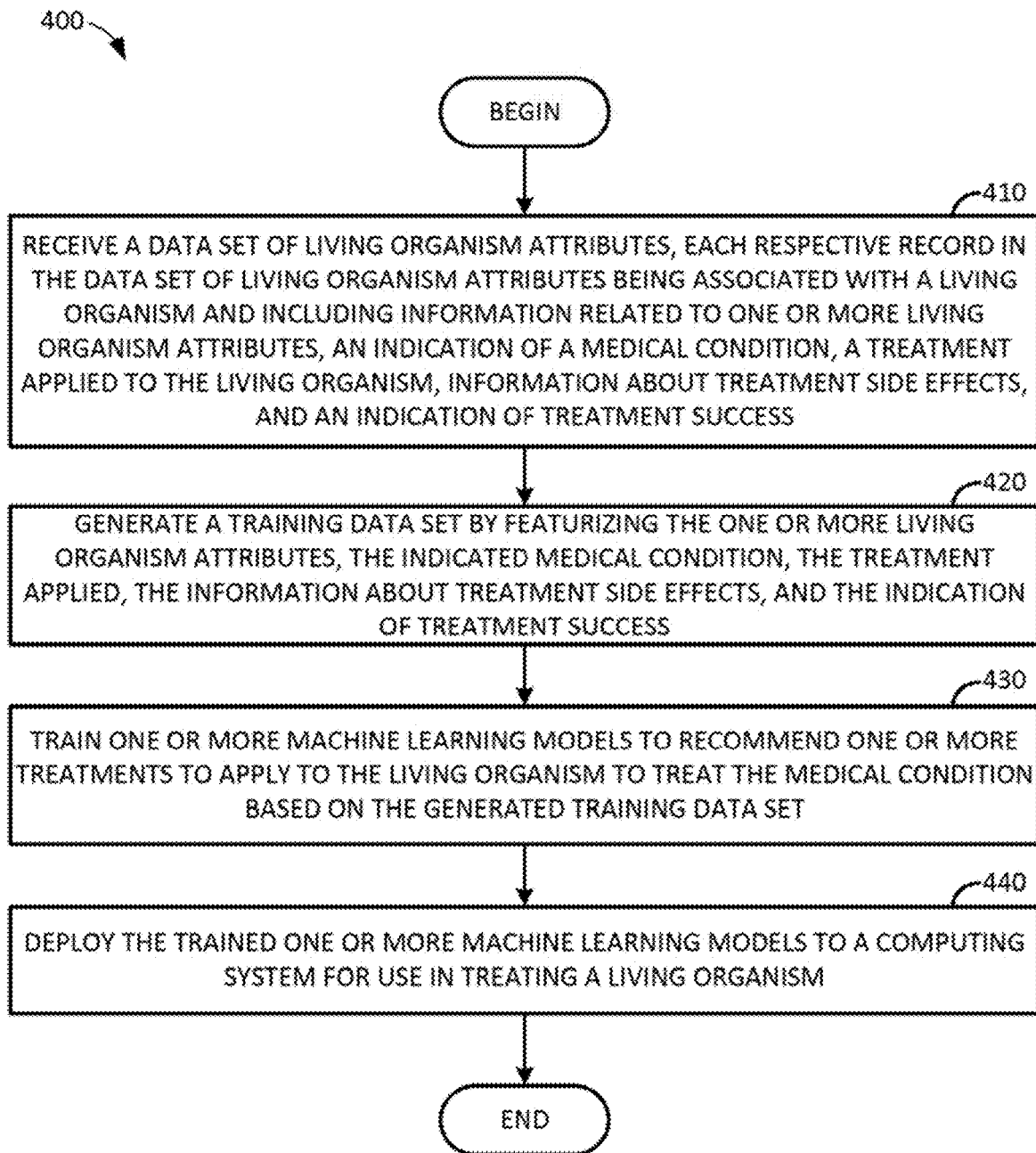


FIG. 4

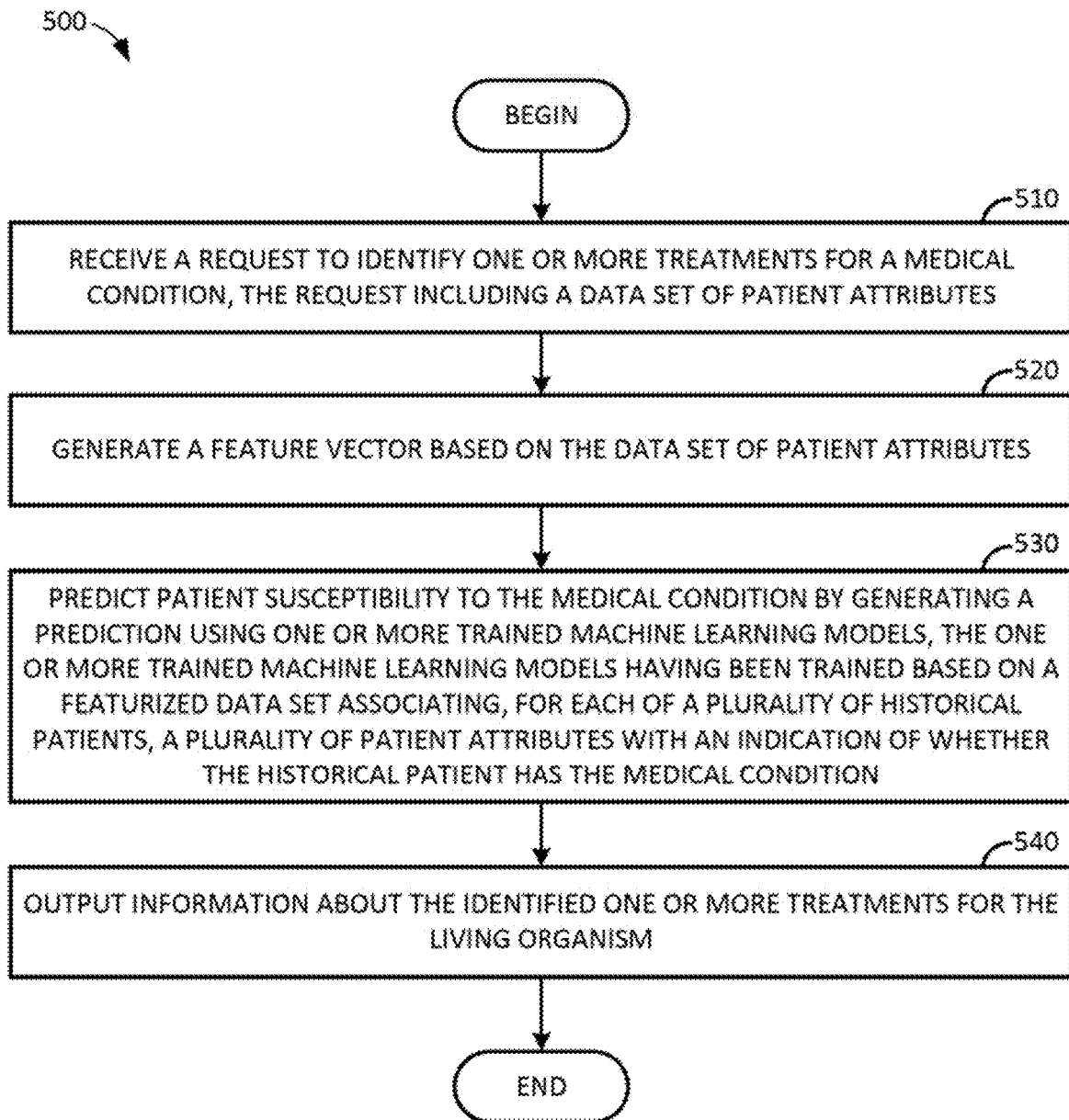


FIG. 5

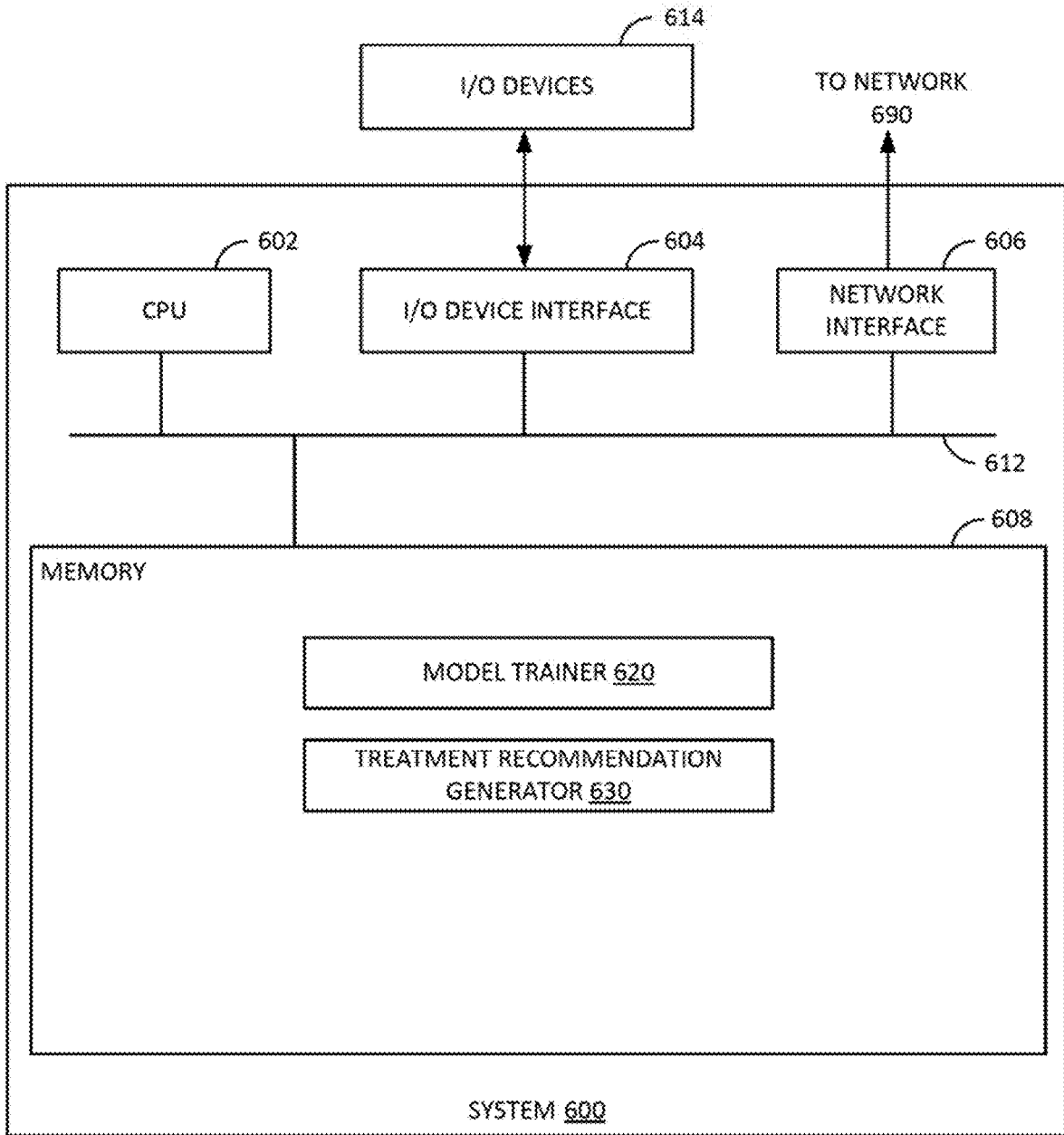


FIG. 6

**RECOMMENDING TREATMENTS TO  
MITIGATE MEDICAL CONDITIONS AND  
PROMOTE SURVIVAL OF LIVING  
ORGANISMS USING MACHINE LEARNING  
MODELS**

CROSS-REFERENCE TO RELATED  
APPLICATIONS

**[0001]** This application claims benefit to U.S. Provisional Patent Application Ser. No. 63/005,916, entitled “Survivability of Illnesses (Viruses, Bacterial Infections, Cancers) Based on Genetic Markers with Correlation to Treatment Strategies,” filed Apr. 6, 2020, and assigned to the assignee hereof, the contents of which are hereby incorporated by reference in its entirety.

BACKGROUND

Field

**[0002]** Embodiments of the present disclosure generally relate to methods for analyzing survivability of illnesses.

Description of the Related Art

**[0003]** Conventional methods for analyzing the survivability of illnesses are generally qualitative and not quantitative.

**[0004]** Therefore, there is a need in the art for more accurate analysis of the survivability of illnesses.

SUMMARY

**[0005]** Embodiments of the present disclosure generally relate to methods for analyzing survivability of illnesses, such as COVID-19. More particularly, embodiments of the present disclosure relate to methods for identifying survivability of illnesses based on genetic markers and other available data.

BRIEF DESCRIPTION OF THE DRAWINGS

**[0006]** So that the manner in which the above recited features of the present disclosure can be understood in detail, a more particular description of the disclosure, briefly summarized above, may be had by reference to embodiments, some of which are illustrated in the appended drawings. It is to be noted, however, that the appended drawings illustrate only exemplary embodiments and are therefore not to be considered limiting of its scope. The disclosure may admit to other equally effective embodiments.

**[0007]** FIGS. 1-3 illustrate flow charts of a method according to embodiments of the present disclosure.

**[0008]** FIG. 4 illustrates example operations that may be performed by a computing system to train one or more machine learning models to recommend treatments for a medical condition based on living organism attributes to promote survivability of the living organism, according to embodiments of the present disclosure.

**[0009]** FIG. 5 illustrates example operations that may be performed by a computing system to identify treatments to a medical condition for a living organism using one or more trained machine learning models, according to embodiments of the present disclosure.

**[0010]** FIG. 6 illustrates an example system in which embodiments of the present disclosure may be implemented.

**[0011]** To facilitate understanding, identical reference numerals have been used, where possible, to designate identical elements that are common to the figures. It is contemplated that elements and features of one embodiment may be beneficially incorporated in other embodiments without further recitation.

DETAILED DESCRIPTION

**[0012]** Embodiments of the present disclosure generally relate to methods for analyzing survivability of illnesses, such as COVID-19. More particularly, embodiments of the present disclosure relate to methods for identifying correlations and influencing factors between genetic markers, lifestyle, and other available data that lead to predictions of the effectiveness of medical treatments, predicting results of mass exposure to an illness based on a population’s genomes and other available data, and providing indicators and methods of visualization for survivability of a viral infection or cancer in any living organism.

Definitions

**[0013]** As used herein, “living organism” refers to any human, animal, plant or other organism that is living or was considered alive at any point.

**[0014]** As used herein, “illness” refers to viruses, bacterial infections, and cancers.

Description

**[0015]** FIG. 1 illustrates a flow chart of a method **100** according to embodiments of the present disclosure. The method **100** generally includes collecting data, standardizing the collected data, generating a testing data set and a training data set, building a correlative model using machine learning, and providing quantitative predictions regarding survivability of illness using the correlative model. As shown in FIG. 2, the testing data set is a subset of data used to validate the model. As shown in FIG. 1, the method **100** includes generating quantitative predictions, examples of which shown in FIGS. 2 and 3.

**[0016]** The minimum data required includes (1) classification of outcome(s) from an illness, such as carrier, non-symptomatic, mild/non-critical (no hospitalization), non-critical (hospitalization), critical or intensive care (hospitalization), death, or other, and/or (2) symptom details from an illness.

**[0017]** Other data includes, but is not limited to, all past, current, and future medical test results, DNA analysis, virus type/taxonomic classification, demographics (age, ethnicity, eye color, skin color, hair color, etc.), climate/location (location (ZIP/postal code)), date (current average weather, seasonality), environmental data (prescriptions, lung capacity, smoker, alcohol consumption, etc.), prior medical procedures and conditions (cancer, high blood pressure, diabetes, pneumonia, bronchitis, hay fever/asthma, viral infections/other diseases (COVID, mumps, measles, chick pox, malaria, lupus or other autoimmune disease), vaccinations (MMR, flu shot, etc.)), blood tests (antibodies, blood type, plasma, O<sub>2</sub> volume, lactic dehydrogenase (LDH), lymphocyte, high-sensitivity C-reactive protein (hsCRP)), diet (standard American diet, ketogenic, carnivore, vegetarian, vegan, kosher, halal, etc.), supplements and naturopathic treatments, work hazard exposures (asbestos, dust), treatments performed (ventilator), other tests per-



formed (CT scan, IR scan, etc.), personal information (lung volume, CO<sub>2</sub>/O<sub>2</sub> exchange volume, sleep pattern, weight, BMI, blood pressure, etc.), and positional/location tracking (GPS, Bluetooth, including PEPP-PT, the Pan-European Privacy-Preserving Proximity Tracing).

**[0018]** At the conclusion of the method **100**, certain quantitative output will be generated. Governmental or healthcare professionals, corporations, or other individuals may then use the quantitative output to undertake a survivability assessment.

**[0019]** In certain embodiments, the quantitative data that is output from method **100** beneficially allows healthcare providers to determine which treatments will be most effective. For example, in certain embodiments, the output quantitative data indicates the percent likelihood that a certain vaccination or other treatment will treat an illness, such as COVID-19, by providing a percent value for the number of individuals who are likely to recover from the illness after receiving the treatment.

**[0020]** Moreover, the quantitative data output from method **100** allows governments, corporations, or individuals to determine where to invest in technology depending on which treatments are more effective.

**[0021]** Embodiments of the present disclosure advantageously replace qualitative conjecture with quantitative evidence, utilizing data science to model the complex relationships as it pertains to illnesses. Embodiments of the present disclosure may be used by individuals to identify their own risks or by doctors, corporations or governments to service or identify exposure for outbreaks that may affect any being to predict the effectiveness of medical treatments, results of mass exposure to an illness based on a population's genomes and other available data, and survivability of a viral infection or cancer in any living organism.

#### EXAMPLE IDENTIFICATION AND RECOMMENDATION OF TREATMENTS FOR MEDICAL CONDITIONS BASED ON MACHINE LEARNING MODELS

**[0022]** Patients who are seeking treatments for a medical condition may have different attributes that may affect the efficacy of a treatment for a medical condition, such as illnesses caused by pathogens (e.g., severe respiratory distress syndrome caused by the SARS-COV-2 virus), cancer, auto-immune disorders, or other causes, based on various factors. For example, patients with a certain blood type may be more responsive to certain types of treatments than patients with other blood types. In another example, treatment efficacy for patients who are overweight and/or do not perform a large amount of physical activity may be significantly different from treatment efficacy for the same treatment for patients who are not clinically overweight and/or are physically active. In still further examples, other attributes, such as exposure to different chemical compounds, or the like, may affect a patient's response to a treatment for a medical condition.

**[0023]** To improve the health of a patient and prevent or mitigate the effects of these various medical conditions, physicians may recommend various treatments to attempt to cure, or at least mitigate the effects of, a medical condition. These actions may include prescribing various medications (which may be more or less effective for different types of patients based on the unique attributes associated with these patients), recommending foods or activities to seek or avoid,

recommending minimization of exposure to certain chemical agents, and the like. Generally, these actions may be recommended by a physician based on generically applicable principles, which may cause recommendations to be made that may not be optimal for any given patient. Further, physicians may make recommendations generally, even though these recommendations may be more relevant for patients having higher susceptibility to a medical condition than patients with lower susceptibility to the medical condition.

**[0024]** Similarly, caretakers of other living organisms may also be interested in identifying treatments to medical conditions that the living organisms under their care are afflicted by. These living organisms may also have unique attributes that make them more or less susceptible to a medical condition and that can affect the efficacy of various treatments for the medical condition. Further, these attributes may also affect the likelihood that the living organism will experience side effects and the severity of those side effects.

**[0025]** Aspects of the present disclosure provide machine learning techniques that allow for the efficacy and severity of side effects of various treatments to be predicted for a living organism, which in turn may be used to recommend treatments for the living organism. By using machine learning models to predict the efficacy and severity of side effects for a living organism based on various attributes of the living organism, aspects of the present disclosure may allow for more accurate targeting of medical interventions. Treatments that are more likely to cure or mitigate the effects of a medical condition may be identified and implemented over treatments that are less likely to cure or mitigate the effects of the medical condition, which may promote survivability of the living organism (e.g., by prioritizing and implementing treatments that are likely to be successful, and skipping or otherwise deprioritizing treatments that are less likely to be effective and/or have a high likelihood of causing severe side effects).

**[0026]** FIG. 4 illustrates example operations for training a machine learning model to recommend treatments for a medical condition experienced by a living organism, according to certain aspects described herein.

**[0027]** As illustrated, operations **400** may begin at block **410**, where a computing system receives a data set of living organism attributes. The data set of living organism attributes may include a plurality of records, and each record in the data set may be associated with a specific living organism and may include information related to one or more living organism attributes, an indication of a medical condition that the living organism has, information about a treatment applied to the living organism, information about side effects of the treatment (e.g., types of side effects and the severity of those side effects), and an indication of treatment success.

**[0028]** In some aspects, the data set of living organism attributes may be received from a plurality of data sources and may be aggregated into a unified data set prior to training one or more machine learning models. The plurality of data sources may include, for example, a secure medical records repository (e.g., a repository of patient medical records subject to the privacy and security requirements of the Health Insurance Portability and Accountability Act or other relevant data privacy regulations) and from one or more other external data sources, such as activity trackers, patient surveys, exposure counters, wearable medical

devices, or the like. Generally, to aggregate the data into the unified data set, information from each of a plurality of sources can be mapped to one or more attributes in the unified data set into which the information is to be mapped, and the appropriate values may be filled into the attributes in the unified data set from the appropriate data source.

**[0029]** The attributes included in each record in the received data set may include a variety of medical, activity, environmental, and other information about or received from the living organism associated with the record. The medical information may include, for example, information such as blood type, blood pressure, known conditions that the patient has, prior surgical history, prescription medications that the user is taking (including, but not limited to, the trade name or active ingredient(s) of the medication and dosage information), family medical history, habits, and the like. The activity information may include, for example, an average number of calories burned per day, an amount of time the patient spent exercising, and the like. Environmental information may include, for example, indications of whether the user has been exposed to or is regularly exposed to various chemicals or types of radiation, the amount of exposure, and/or other environmental information that may influence susceptibility to a medical condition, efficacy of treatments for the medical condition, and/or likelihood of experiencing severe side effects from the treatments.

**[0030]** At block 420, the computing system generates a training data set by featurizing the one or more living organism attributes, the indication of the medical condition that the living organism has, the information about the treatment applied to the living organism, the information about side effects of the treatment (e.g., types of side effects and the severity of those side effects), and the indication of treatment success. To featurize the one or more living organism attributes, the raw data in each record in the received data set may be transformed into machine-readable or machine-usable data that can be used to train a machine learning model. Generally, raw data may be transformed into numerical data representing, for example, a binary choice (e.g., whether a patient is associated with a given attribute or is not associated with the given attribute, such as whether a living organism has a given medical condition, is taking a given medication, or the like), one of a plurality of categories (e.g., where an attribute has a range of values, and different sub-ranges are probative of different levels of susceptibility to a medical condition, such as ranges of weight, ranges of exposure, etc.), or numerical data scaled based on a scaling factor.

**[0031]** The computing system can use one or more predefined rules to determine how to featurize each of the one or more living organism attributes. Each attribute to be included in a training data set may be associated with a rule indicating how the underlying raw data from the received data set is to be transformed into a feature usable in training a machine learning model to predict susceptibility to a medical condition. In some aspects, the rules may define how multiple related data items may be aggregated into a single value, and the single value may be featurized. In another example, multiple different values may map to a same featurized value. For example, if an attribute is whether a patient has been prescribed or is otherwise taking over-the-counter allergy medication (e.g., a binary feature), it may be recognized that there are many types of allergy medications that a patient can be taking. Thus, the rule may

recognize that if the data set includes information indicating that the patient is taking one of the various types of allergy medications, regardless of the exact active ingredient or form of administration. In another example, the rules may define upper and lower bound values for classification of an attribute into one of a plurality of categories. For example, given patient weight and height as information included in a record in the received data set, an attribute may be defined as the patient's body mass index (BMI), and different values may be assigned to the attribute based on different BMI ranges (e.g., where a first value corresponds to underweight BMIs, a second value corresponds to normal BMIs, a third value corresponds to overweight BMIs, and a fourth value corresponds to obese BMIs).

**[0032]** In some aspects, some attributes may be determined based on raw data, and the one or more predefined rules may specify a scaling factor associated with the devices that recorded the raw data to use in scaling the data (e.g., prior to featurization). The scaling factor may be, for example, associated with an accuracy of a measurement device, which may be defined a priori according to manufacturer specifications or prior experience with the measurement device. For example, where an attribute includes a size of an anatomical feature captured using one or more imaging devices (e.g., X-ray machines, magnetic resonance imagery machines, computed tomography (CT) machines, etc.), the raw size information may be adjusted based on an expected measurement error for the source imaging device. If, for example, an imaging device is known to be accurate to within  $n$  percent, the raw data may be scaled to a value of  $100+n$  percent or  $100-n$  percent, depending on the specific direction of error, developer choice, or the like. The scaled value may be preserved as the value associated with an attribute or may be further featurized into a binary feature or a feature with a fixed set of values, as discussed above.

**[0033]** In some aspects, information about the side effects may be structured as a collection of side effect vectors, with each vector including an identification of a side effect and an indication of a severity of the side effect. Generally, the identified side effects may be grouped into different classifications of side effects, such as side effects related to the gastrointestinal tract, cardiovascular system, nervous system, etc. The indication of the severity of the side effect may be selected from a plurality of categories, from no side effects to death. Rules may define a base severity category into which a side effect is to be assigned, as the occurrence of certain side effects may be considered more serious than the occurrence of other side effects. For example, a side effect of temporary paralysis may be mapped to a severe side effect category according to an a priori defined rule.

**[0034]** In some aspects, the attributes included in the received data set may be reduced based on various filtering or selection techniques. It may be noticed, for example, that records associated with multiple living organisms include similar values for a particular attribute, regardless of whether the living organism has the medical condition. Because values for the particular attribute are similar for disparate outcomes across records in the data set, it may be determined that the attribute is not probative of whether a living organism is susceptible to the medical condition, whether a treatment for the medical condition will be effective for a living organism, and/or whether a living organism will experience severe side effects from the treatment. Thus, the attribute may be removed from each of the

records in the data set, which may reduce the amount of data processed while training the machine learning models. In another example, statistical tests can be used to determine whether an attribute is independent or dependent by using techniques such as chi-squared testing to determine whether observations deviate from an expected outcome for a particular analysis. In still further examples, various machine learning techniques can be used to assign an importance or significance value to each attribute. Attributes in the received data set having importance or significance values exceeding a threshold value may be retained in the received data set, while attributes having importance or significance values below the threshold value may be removed from the received data set.

**[0035]** In some aspects, the data set may not include a value for an attribute for a given living organism. To allow for each of the records in the data set to have a same number of attributes, the record for that given living organism may be modified with a value for the attribute indicating that the attribute does not apply to the living organism. For example, the value for the attribute may be a reserved value (e.g., a predefined magic number), a null value, or the like.

**[0036]** At block 430, the computing system trains one or more machine learning models to recommend one or more treatments to apply to a living organism to treat the medical condition based on the generated training data set. The one or more machine learning models may be various types of machine learning models configured to generate various outputs. For example, the machine learning models may include one or more of probabilistic models, neural networks, clustering models, or other appropriate machine learning models. Generally, a probabilistic model may be configured to generate a probability distribution over a plurality of treatment options, where the probability value associated the treatment option corresponds to a likelihood that the treatment option is effective for living organisms with the given set of attributes. In some aspects, a first machine learning model may be trained to predict a likelihood of treatment success, and a second machine learning model may be trained to predict a likelihood of the living organism experiencing different levels of side effects for each treatment. A clustering algorithm may be used to identify living organisms having similar attributes to a given living organism whose attributes are received as input. Information about the identified living organisms can then be used, as discussed in further detail below, to identify recommended treatments for the medical condition. For example, recommended treatments for the medical condition may be identified based on ratios of historical living organisms in a set of similar living organisms being treated using a particular treatment to the total number of historical living organisms in the set of similar living organisms.

**[0037]** At block 440, the computing system deploys the trained one or more machine learning models to one or more other computing systems for use in treating a living organism. As discussed in further detail below, these computing systems can use the trained machine learning models to identify treatments that have been applied to similar living organisms. Based on the identification of these treatments, the computing system can predict the efficacy of each treatment and a likelihood that the living organism will experience severe side effects from the treatment and use these predictions to identify treatments that are likely to cure or mitigate the effects of a medical condition experienced by a living

organism while minimizing the likelihood of the living organism experiencing severe side effects from the treatment.

**[0038]** FIG. 5 illustrates example operations 500 that may be performed by a computing system to identify and/or recommend treatments for a medical condition based on one or more machine learning models.

**[0039]** As illustrated, operations 500 may begin at block 510, where the computing system receives a request to identify one or more treatments for a medical condition. The request generally includes a raw data set of living organism attributes and information about a medical condition for which the living organism to be treated. Like the records discussed above with respect to a data set used to train the one or more machine learning models, the raw data set of living organism attributes may include information from a secure medical records repository and from one or more other external data sources, such as activity trackers, patient surveys, exposure counters, wearable medical devices, or the like.

**[0040]** The attributes included in request may include a variety of medical, activity, environmental, and other information about or received from the living organism associated with the record. The medical information may include, for example, information such as blood type, blood pressure, known conditions that the living organism has, prior surgical history, prescription medications that the living organism is taking (including, but not limited to, the trade name or active ingredient(s) of the medication and dosage information), family medical history, habits, and the like. The activity information may include, for example, an average number of calories burned per day, an average amount of time spent exercising, and the like. Environmental information may include, for example, indications of whether the living organism has been exposed to or is regularly exposed to various chemicals or types of radiation, the amount of exposure, and/or other environmental information that may influence susceptibility to a medical condition, efficacy of treatments for the medical condition, and/or a likelihood of experiencing severe side effects from a treatment.

**[0041]** At block 520, the computing system generates a feature vector based on the data set of living organism attributes. As discussed, to generate the feature vector, the computing system can transform the raw data in the request into machine-readable or machine-usable data that can be used as input into a trained machine learning model. Generally, raw data may be transformed into numerical data representing, for example, a binary choice (e.g., whether a living organism is associated with a given attribute or is not associated with the given attribute, such as whether a living organism has a given medical condition, is taking a given medication, or the like), one of a plurality of categories (e.g., where an attribute has a range of values, and different sub-ranges are probative of different levels of susceptibility to a medical condition, such as ranges of weight, ranges of exposure, etc.), or numerical data scaled based on a scaling factor.

**[0042]** The computing system can use one or more predefined rules to determine how to featurize each of the one or more living organism attributes. Each attribute to be used in predicting efficacy of a treatment for a medical condition and/or a likelihood of experiencing severe side effects from the treatment may be associated with a rule indicating how the underlying raw data from the received data set is to be

transformed into a feature usable by a machine learning model to predict efficacy of a treatment for a medical condition and/or a likelihood of experiencing severe side effects from the treatment. In some aspects, the rules may define how multiple related data items may be aggregated into a single value, and the single value may be featurized. In another example, multiple different values may map to a same featurized value. In another example, the rules may define upper and lower bound values for classification of an attribute into one of a plurality of categories.

**[0043]** In some aspects, some attributes may be determined based on raw data, and the one or more predefined rules may specify a scaling factor associated with the devices that recorded the raw data to use in scaling the data (e.g., prior to featurization). The scaling factor may be, for example, associated with an accuracy of a measurement device, which may be defined a priori according to manufacturer specifications or prior experience with the measurement device. The scaled value may be preserved as the value associated with an attribute or may be further featurized into a binary feature or a feature with a fixed set of values, as discussed above.

**[0044]** In some aspects, the attributes included in the request may be reduced based on various filtering or selection techniques. The filtering or selection techniques may be defined based on the filtering or selection techniques used to filter data in a training data set used to train the one or more machine learning models. To reduce the information included in the feature vector down to a minimal set of information needed for the one or more machine learning models to predict efficacy of a treatment for a medical condition and/or a likelihood of experiencing severe side effects from the treatment, attributes that are known a priori to not be probative of whether someone is susceptible to the medical condition may be removed from the data set included in the request.

**[0045]** In some aspects, the data set may not include a value for an attribute. To allow for the feature vector to have a same number of attributes as the records in the training data set used to train the one or more machine learning models, the feature vector may be modified with a value for the attribute indicating that the attribute does not apply to the living organism. For example, the value for the attribute may be a reserved value (e.g., a predefined magic number), a null value, or the like.

**[0046]** At block 530, the computing system identifies one or more recommended treatments by generating a prediction using one or more trained machine learning models. As discussed above, the machine learning models may have been previously trained based on a featurized data set associating, for each historical living organism of a plurality of historical living organisms, a plurality of attributes in history for the historical living organism with an indication of whether the historical living organism has the medical condition.

**[0047]** In some aspects, the one or more machine learning models may include probabilistic models that are trained to output, for a given input, a probability distribution over a universe of possible outcomes. In some aspects, the probability distribution may be generated over each of the treatments for which data exists in a training data set, with the probability value associated with each treatment serving as a proxy for a likelihood of efficacy for treating the medical condition. In some aspects, multiple probabilistic models

can be used to predict which treatments are likely to be effective for the living organism, and each model of the multiple probabilistic models may be associated with a weighting value. A score serving as a proxy for efficacy of a treatment for the medical condition may be calculated as a weighted average of the probability scores output by each of the multiple probabilistic models.

**[0048]** In some aspects, the one or more machine learning models may also or alternatively include one or more clustering models that are trained to identify a set of matching historical living organisms having similar data sets of attributes. To identify recommended treatments for the medical condition, a score can be generated based the treatment efficacy metrics associated with different treatments used for the living organisms in the set of matching historical living organisms who are identified as having the medical condition. For example, a score may be generated based on a weighted average of the treatment efficacy metrics for each treatment. In some aspects, a score for each treatment may also be adjusted based on the severity of side effects experienced by the living organisms that are treated using that specific treatment; in such a case, a metric related to the severity of side effects may be used, for example, as a scaling factor, where the efficacy of a treatment is scaled downwards to account for the severity of the side effects. By doing so, the system can decline to recommend, for a living organism having a given set of attributes, treatments that have high efficacy metrics but are likely to lead to severe side effects.

**[0049]** In some aspects, a probabilistic model and a clustering model (as well as other machine learning models) may be used in conjunction with each other to predict efficacy of a treatment for a medical condition and/or a likelihood of experiencing severe side effects from the treatment. In one example, a probabilistic model may be associated with a first weighting value, and the clustering model may be associated with a second weighting value. The probability score—representing efficacy of a treatment for a medical condition and/or a likelihood of experiencing severe side effects from the treatment—may be calculated as sum of the score generated by the probabilistic model, weighted by the first weighting value, and the score generated by the clustering model, weighted by the second weighting value.

**[0050]** At block 540, the computing system outputs the identified one or more treatments for the living organism. The identified treatments for the living organism may include information about one or more treatments that may be relevant for the living organism as well as information about potential side effects of the treatment and likelihood of occurrence based on statistics calculated from similar living organisms who received each treatment. In some aspects, the identified treatments for the living organism may be output as an ordered list of treatments, with higher scoring treatments (e.g., treatments with high predicted efficacy and low predicted incidence of severe side effects) being at the top of the ordered list, and lower scoring treatments (e.g., treatments with low predicted efficacy or treatments with high predicted efficacy and high predicted incidence of severe side effects) being at the bottom of the ordered list.

#### EXAMPLE SYSTEMS FOR IDENTIFYING AND/OR RECOMMENDING TREATMENTS FOR A MEDICAL CONDITION USING MACHINE LEARNING MODELS

**[0051]** FIG. 6 illustrates an example system 600 that can train and use machine learning models to identify and/or

recommend treatments for a medical condition, according to certain embodiments described herein.

[0052] As shown, system 600 includes a central processing unit (CPU) 602, one or more I/O device interfaces 604 that may allow for the connection of various I/O devices 614 (e.g., keyboards, displays, mouse devices, pen input, etc.) to the system 600, network interface 606 through which system 600 is connected to network 660 (which may be a local network, an intranet, the internet, or any other group of computing devices communicatively connected to each other), a memory 608, storage 610, and an interconnect 612.

[0053] CPU 602 may retrieve and execute programming instructions stored in the memory 608. Similarly, the CPU 602 may retrieve and store application data residing in the memory 608. The interconnect 612 transmits programming instructions and application data, among the CPU 602, I/O device interface 604, network interface 604, memory 608, and storage 610.

[0054] CPU 602 is included to be representative of a single CPU, multiple CPUs, a single CPU having multiple processing cores, and the like.

[0055] Memory 608 is representative of a volatile memory, such as a random access memory, or a nonvolatile memory, such as nonvolatile random access memory, phase change random access memory, or the like. As shown, memory 608 includes a model trainer 620 and a treatment recommendation generator 630.

[0056] Model trainer 620 may be configured to perform the operations discussed herein (e.g., with respect to operations 400 illustrated in FIG. 4 and/or other operations) to train and deploy one or more machine learning models for recommending treatments for a medical condition based on living organism attributes. As discussed, model trainer 620 can receive data from a plurality of data sources (including, but not limited to, a secure medical records data source, a physical activity records data source, a medicine usage data source, and/or other data sources in which attributes that may be predictive, alone or in isolation, of efficacy of a treatment for a medical condition and/or a likelihood of experiencing severe side effects from the treatment for a living organism may be stored) and generate a training data set by featurizing the one or more attributes. Model trainer 620 may be configured to train one or more machine learning models based on the generated training data set. As discussed, the one or more machine learning models may include probabilistic models, clustering-based models, and/or other machine learning models that may be used to recommend treatments for a medical condition based on predicted efficacy and severity of side effects for a living organism having some given input of a plurality of living organism attributes. Model trainer 620 may then deploy the trained one or more machine learning models for use (e.g., to treatment recommendation generator 630 and/or one or more external computing systems accessible via network 660).

[0057] Treatment recommendation generator 630 may be configured to perform the operations discussed herein (e.g., with respect to operations 500 illustrated in FIG. 5 and/or other operations) to identify potential treatments for a medical condition based on one or more machine learning models and living organism attributes. As discussed, treatment recommendation generator 630 may use the one or more machine learning models trained by model trainer 620 to identify treatments for a medical condition having high

probabilities of efficacy and lower likelihood of severe side effects. To do so, treatment recommendation generator 630 can receive a request including a data set of living organism attributes and generate a feature vector based on the data set of living organism attributes. The feature vector may be provided as input into one or more machine learning models to generate a score for each of a plurality of treatments for a medical condition. Based on the generated scores, treatment recommendation generator 630 can identify one or more treatments that are candidates for the living organism. These treatments may generally be treatments having a high probability of efficacy for the living organism, given the living organism's attributes, with a low probability of severe side effects (which, as discussed above, may be used to scale the predicted efficacy so that treatments are effectively penalized for a high likelihood of severe side effects).

#### ADDITIONAL CONSIDERATIONS

[0058] The preceding description is provided to enable any person skilled in the art to practice the various embodiments described herein. Various modifications to these embodiments will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other embodiments. For example, changes may be made in the function and arrangement of elements discussed without departing from the scope of the disclosure. Various examples may omit, substitute, or add various procedures or components as appropriate. Also, features described with respect to some examples may be combined in some other examples. For example, an apparatus may be implemented or a method may be practiced using any number of the embodiments set forth herein. In addition, the scope of the disclosure is intended to cover such an apparatus or method that is practiced using other structure, functionality, or structure and functionality in addition to, or other than, the various embodiments of the disclosure set forth herein. It should be understood that any aspect of the disclosure disclosed herein may be embodied by one or more elements of a claim.

[0059] As used herein, a phrase referring to "at least one of" a list of items refers to any combination of those items, including single members. As an example, "at least one of: a, b, or c" is intended to cover a, b, c, a-b, a-c, b-c, and a-b-c, as well as any combination with multiples of the same element (e.g., a-a, a-a-a, a-a-b, a-a-c, a-b-b, a-c-c, b-b, b-b-b, b-b-c, c-c, and c-c-c or any other ordering of a, b, and c).

[0060] As used herein, the term "determining" encompasses a wide variety of actions. For example, "determining" may include calculating, computing, processing, deriving, investigating, looking up (e.g., looking up in a table, a database or another data structure), ascertaining and the like. Also, "determining" may include receiving (e.g., receiving information), accessing (e.g., accessing data in a memory) and the like. Also, "determining" may include resolving, selecting, choosing, establishing and the like.

[0061] The methods disclosed herein comprise one or more steps or actions for achieving the methods. The method steps and/or actions may be interchanged with one another without departing from the scope of the claims. In other words, unless a specific order of steps or actions is specified, the order and/or use of specific steps and/or actions may be modified without departing from the scope of the claims. Further, the various operations of methods described above may be performed by any suitable means capable of per-

forming the corresponding functions. The means may include various hardware and/or software component(s) and/or module(s), including, but not limited to a circuit, an application specific integrated circuit (ASIC), or processor. Generally, where there are operations illustrated in figures, those operations may have corresponding counterpart means-plus-function components with similar numbering.

**[0062]** The various illustrative logical blocks, modules and circuits described in connection with the present disclosure may be implemented or performed with a general purpose processor, a digital signal processor (DSP), an application specific integrated circuit (ASIC), a field programmable gate array (FPGA) or other programmable logic device (PLD), discrete gate or transistor logic, discrete hardware components, or any combination thereof designed to perform the functions described herein. A general-purpose processor may be a microprocessor, but in the alternative, the processor may be any commercially available processor, controller, microcontroller, or state machine. A processor may also be implemented as a combination of computing devices, e.g., a combination of a DSP and a microprocessor, a plurality of microprocessors, one or more microprocessors in conjunction with a DSP core, or any other such configuration.

**[0063]** A processing system may be implemented with a bus architecture. The bus may include any number of interconnecting buses and bridges depending on the specific application of the processing system and the overall design constraints. The bus may link together various circuits including a processor, machine-readable media, and input/output devices, among others. A user interface (e.g., keypad, display, mouse, joystick, etc.) may also be connected to the bus. The bus may also link various other circuits such as timing sources, peripherals, voltage regulators, power management circuits, and the like, which are well known in the art, and therefore, will not be described any further. The processor may be implemented with one or more general-purpose and/or special-purpose processors. Examples include microprocessors, microcontrollers, DSP processors, and other circuitry that can execute software. Those skilled in the art will recognize how best to implement the described functionality for the processing system depending on the particular application and the overall design constraints imposed on the overall system.

**[0064]** If implemented in software, the functions may be stored or transmitted over as one or more instructions or code on a computer-readable medium. Software shall be construed broadly to mean instructions, data, or any combination thereof, whether referred to as software, firmware, middleware, microcode, hardware description language, or otherwise. Computer-readable media include both computer storage media and communication media, such as any medium that facilitates transfer of a computer program from one place to another. The processor may be responsible for managing the bus and general processing, including the execution of software modules stored on the computer-readable storage media. A computer-readable storage medium may be coupled to a processor such that the processor can read information from, and write information to, the storage medium. In the alternative, the storage medium may be integral to the processor. By way of example, the computer-readable media may include a transmission line, a carrier wave modulated by data, and/or a computer readable storage medium with instructions stored

thereon separate from the wireless node, all of which may be accessed by the processor through the bus interface. Alternatively, or in addition, the computer-readable media, or any portion thereof, may be integrated into the processor, such as the case may be with cache and/or general register files. Examples of machine-readable storage media may include, by way of example, RAM (Random Access Memory), flash memory, ROM (Read Only Memory), PROM (Programmable Read-Only Memory), EPROM (Erasable Programmable Read-Only Memory), EEPROM (Electrically Erasable Programmable Read-Only Memory), registers, magnetic disks, optical disks, hard drives, or any other suitable storage medium, or any combination thereof. The machine-readable media may be embodied in a computer-program product.

**[0065]** A software module may comprise a single instruction, or many instructions, and may be distributed over several different code segments, among different programs, and across multiple storage media. The computer-readable media may comprise a number of software modules. The software modules include instructions that, when executed by an apparatus such as a processor, cause the processing system to perform various functions. The software modules may include a transmission module and a receiving module. Each software module may reside in a single storage device or be distributed across multiple storage devices. By way of example, a software module may be loaded into RAM from a hard drive when a triggering event occurs. During execution of the software module, the processor may load some of the instructions into cache to increase access speed. One or more cache lines may then be loaded into a general register file for execution by the processor. When referring to the functionality of a software module, it will be understood that such functionality is implemented by the processor when executing instructions from that software module.

**[0066]** The following claims are not intended to be limited to the embodiments shown herein, but are to be accorded the full scope consistent with the language of the claims. Within a claim, reference to an element in the singular is not intended to mean "one and only one" unless specifically so stated, but rather "one or more." Unless specifically stated otherwise, the term "some" refers to one or more. No claim element is to be construed under the provisions of 35 U.S.C. § 112(f) unless the element is expressly recited using the phrase "means for" or, in the case of a method claim, the element is recited using the phrase "step for." All structural and functional equivalents to the elements of the various embodiments described throughout this disclosure that are known or later come to be known to those of ordinary skill in the art are expressly incorporated herein by reference and are intended to be encompassed by the claims. Moreover, nothing disclosed herein is intended to be dedicated to the public regardless of whether such disclosure is explicitly recited in the claims.

1. A method for training machine learning models to recommend treatments for a living organism to address a medical condition, comprising:

receiving a data set of attributes, each respective record in the data set of attributes being associated with a living organism and including information related to one or more attributes, an indication of a medical condition, a treatment applied to the living organism, information about side effects of the treatment and a severity of the side effects, and an indication of treatment success;

- generating a training data set by featurizing the one or more attributes, the indicated medical condition, the treatment applied, the information about side effects of the treatment and the severity of the side effects, and the indication of treatment success;
- training one or more machine learning models to recommend one or more treatments to apply to the living organism to treat the medical condition based on the generated training data set; and
- deploying the trained one or more machine learning models to a computing system for use in treating a living organism.
2. The method of claim 1, wherein featurizing the one or more attributes comprises:
- for each respective medical attribute of the one or more attributes, assigning one of a plurality of values, each value indicating a classification of the respective medical attribute into one of a plurality of categories.
3. The method of claim 1, wherein generating the training data set comprises:
- scaling a value of an item in the data set based on a scaling factor associated with an accuracy of a source from which the value was obtained; and
- featurizing the scaled value of the item.
4. The method of claim 1, further comprising:
- replacing null values for features in the received data set with an indication that the features do not apply to the living organism.
5. The method of claim 1, wherein the data set of attributes is received from a plurality of external data sources.
6. The method of claim 5, further comprising:
- aggregating information from the plurality of external data sources into a single record for each living organism.
7. The method of claim 5, wherein the plurality of external data sources comprises a secure medical records data source and one or more other data sources.
8. The method of claim 7, wherein the one or more other data sources include one or more of a physical activity records data source, or a medicine usage data source.
9. The method of claim 1, wherein the one or more machine learning models comprise clustering-based machine learning models.
10. The method of claim 1, wherein the one or more machine learning models comprise probabilistic models in which efficacy of each of a universe of treatments is represented by a probability distribution over each treatment in the universe of treatments for the medical condition.
11. A method for identifying treatments for a living organism to treat a medical condition based on one or more machine learning models, comprising:
- receiving a request to identify one or more recommended treatments for a medical condition, the request including a data set of living organism attributes;
- generating a feature vector based on the data set of living organism attributes;
- identifying the one or more recommended treatments by generating a prediction using one or more trained machine learning models, the one or more trained machine learning models having been trained based on a featurized data set including, for each historical living organism of a plurality of historical living organisms, one or more attributes, an indication of a medical condition, a treatment applied to the living organism, information about side effects of the treatment and a severity of the side effects, and an indication of treatment success; and
- outputting information about the identified one or more treatments for the living organism.
12. The method of claim 11, wherein the one or more trained machine learning models comprise one or more probabilistic models trained to generate a probability distribution corresponding to a likelihood of each of a plurality of treatments being successful for the living organism having the medical condition and any potential side effects and severity of side effects
13. The method of claim 12, wherein identifying the one or more treatments comprises:
- for each of a plurality of treatments, generating a probability score for the treatment as a weighted average of a likelihood of success generated by each of the one or more trained machine learning models, each model of the one or more trained learning model being associated with a weighting value to assign to a likelihood of the living organism having the medical condition; and
- selecting treatments in the plurality of treatments having a probability score higher than a threshold probability score.
14. The method of claim 11, wherein the one or more trained machine learning models comprise one or more clustering models trained to identify a set of matching historical living organisms of the plurality of historical living organisms having similar data sets of attributes to the living organism.
15. The method of claim 14, wherein identifying the one or more treatments comprises:
- identifying, in the set of matching historical living organisms, a set of treatments applied to living organisms in the set of matching historical living organisms;
- for each treatment of the set of treatments applied to historical living organisms in the set of matching historical living organisms, calculating an average success rate based on success information associated with each historical living organism; and
- selecting treatments from the set of treatments having average success rates exceeding a threshold success rate.
16. The method of claim 11, wherein:
- the one or more trained machine learning models comprise a probabilistic model configured to generate a probability distribution corresponding to a likelihood of each of a plurality of treatments being successful for the living organism having the medical condition and a clustering model configured to identify a set of matching historical living organisms having similar data sets of attributes to the living organism, and
- the one or more recommended treatments are identified based on a weighted average of a probability of success calculated by the probabilistic model and an average success rate for similar living organisms in the set of matching historical living organisms.
17. The method of claim 11, wherein generating the feature vector comprises: for each attribute in the data set, assigning one of a plurality of numerical values for the attribute based on a value of the attribute in the data set, each value indicating a classification of the respective attribute into one of a plurality of categories.

**18.** The method of claim **11**, wherein generating the feature vector comprises:

- scaling a value of an attribute in the data set based on a scaling factor associated with an accuracy of a source from which the value was obtained; and
- featurizing the scaled value of the item.

**19.** The method of claim **11**, wherein generating the feature vector comprises: replacing null values for features in the data set with an indication that the features do not apply to the living organism.

**20.** A system for identifying treatments for living organism to treat a medical condition based on one or more machine learning models, comprising:

- a memory having instructions stored thereon; and
- a processor configured to execute the instructions to cause the system to:
  - receive a request to identify one or more recommended treatments for a medical condition, the request including a data set of living organism attributes;
  - generate a feature vector based on the data set of living organism attributes;

identify the one or more recommended treatments by generating a prediction using one or more trained machine learning models, the one or more trained machine learning models having been trained based on a featurized data set including, for each historical living organism of a plurality of historical living organisms, one or more living organism attributes, an indication of a medical condition, a treatment applied to the living organism, information about side effects of the treatment and a severity of the side effects, and an indication of treatment success; and output information about the identified one or more treatments for the living organism.

**21.** The method of claim **11**, wherein the medical condition comprises respiratory complications caused by SARS-COV-2, and the recommend treatment comprises one or more of vaccination against SARS-COV-2 or use of a ventilator for a patient having respiratory complications caused by SARS-COV-2.

\* \* \* \* \*