

SYSTEMS AND METHODS FOR DYNAMIC RAMAN PROFILING OF BIOLOGICAL DISEASES AND DISORDERS AND FEATURE ENGINEERING METHODS THEREOF**CROSS REFERENCE TO RELATED APPLICATIONS**

[0001] This application claims priority to U.S. Provisional Patent Application Serial No. 63/350,090, filed June 8, 2022, which is hereby incorporated by reference in its entirety.

BACKGROUND

[0002] Dynamic biological responses may be indicative of underlying biological processes having structural and functional significance for humans. For example, aberrant or abnormal dynamic biological response may be associated with many biological conditions, such as diseases and disorders. Examples of such biological conditions may include neurological conditions (*e.g.*, autism spectrum disorder, schizophrenia, or attention-deficit/hyperactivity disorder (ADHD)), neurodegenerative conditions (*e.g.*, amyotrophic lateral sclerosis (ALS), Alzheimer's disease, Parkinson's disease, and Huntington's disease), and cancers (*e.g.*, pediatric cancer).

SUMMARY

[0003] Given the above background, there is a need for accurate methods and systems for the diagnosis of biological conditions, and especially for non-invasive diagnosis. Such diagnosis may be based on accurate profiling of biomarkers detectable with non-invasive methods for diagnosis of the biological conditions. The present disclosure provides improved systems and methods for accurate diagnosis of biological conditions based on analysis of dynamic biological response data from non-invasively obtained biological samples from subjects. Such improved systems and methods for accurate diagnosis of biological conditions may be based on a combination of Raman profiling of biological samples and artificial intelligence data analysis. The present disclosure addresses these needs, for example, by providing a biological sample biomarker for diagnosis of biological conditions. The biological sample includes a human biological specimen that is associated with incremental growth. Such a biological sample could be a hair shaft, a tooth, and a nail. The non-invasive biomarker of the present disclosure can be used for the diagnosis of young children, even infants younger than one year old.

[0004] In an aspect, the present disclosure provides a method for predicting a subject's diagnostic status with respect to disease or disorder of a subject, comprising: (a) exposing a biological sample of the subject to a light source, wherein the biological sample comprises a

tooth sample, a hair sample, or a nail sample; (b) acquiring a plurality of Raman spectra from the exposed biological sample; (c) processing the plurality of Raman spectra to generate a spatial map of the plurality of Raman spectra; and (d) predicting a subject's diagnostic status with respect to a disease or disorder based at least in part on the spatial map of the plurality of Raman spectra. In some embodiments, the light source comprises a laser.

[0005] In some embodiments, the analyzing determines temporal dynamics of underlying biological processes. In some embodiments, the analyzing comprises reducing a dimensionality of the plurality of Raman spectra (*e.g.*, by independent components analysis) prior to the processing. In some embodiments, the optical signal is generated by a light source (*e.g.*, a laser). In some embodiments, the biological sample comprises the tooth sample. In some embodiments, the method further comprises detecting or monitoring changes in a temporal stress profile (*e.g.*, one or more traces) that are indicative of a temporal response of the subject. In some embodiments, the temporal response comprises a biochemical response. In some embodiments, the temporal response comprises a biological response, a physiological response, an anatomical response, a treatment response, a stress-related response, or a combination thereof. In some embodiments, the plurality of Raman spectra comprises from about 200 to about 3700 wave numbers. In some embodiments, the acquiring comprises using Raman spectroscopy microscope. In some embodiments, the Raman spectroscopy microscope comprises an 50X air coupled objective, 63X water immersion coupled objection, or any combination thereof. In some embodiments, the laser comprises a wavelength of about 785 nm, a wavelength of about 532 nm, or any combination thereof. In some embodiments, the acquiring is performed using an integration time of about 0.2 seconds to about 0.3 seconds. In some embodiments, the acquiring comprises moving the biological sample with a step size of about 2 microns to about 5 microns, subsequent to acquiring a Raman spectrum of the plurality of Raman spectra.

[0006] In some embodiments, the disease or disorder comprises autism spectrum disorder (ASD), attention deficit/hyperactivity disorder (ADHD), amyotrophic lateral sclerosis (ALS), schizophrenia, irritable bowel disease (IBD), pediatric kidney disease, kidney transplant rejection, pediatric cancer or any combination thereof. In some embodiments, the disease or disorder comprises the ASD. In some embodiments, the subject is a human. In some embodiments, the subject is an adult. In some embodiments, the subject is between the ages of about 12 and about 5 years old. In some embodiments, the subject is less than about 12, 11, 10, 9, 8, 7, 5, 4, 3, 2, or 1 year(s) old. In some embodiments, the subject is at least about 1, 2, 3, 4, 5, 6,

7, 8, 9, 10, 11, or 12 year(s) old. In some embodiments, at least a portion of the temporal Raman profile corresponds to a prenatal period of the subject.

[0007] In some embodiments, predicting a subject's diagnostic status with respect to a disease or disorder comprises processing the spatial map using a trained model. In some embodiments, the processing comprises extracting features from the spatial map (*e.g.*, by recurrence quantification analysis), and analyzing the features using the trained model. In some embodiments, the processing comprises computational analysis of temporal dynamics derived from the spatial map, *e.g.*, by application of dimensionality reduction techniques, including independent component analysis (ICA) and/or principal component analysis (PCA), followed by the subsequent application of recurrence quantification analysis (RQA) to extract computational features descriptive of the dimensions derived from ICA/PCA.

[0008] In some embodiments, the processing comprises determining one or more features of the temporal dynamics of the one or more traces. In some embodiments, the temporal dynamics of the one or more traces are determined by data analysis methods. In some embodiments, the data analysis methods apply one or more of the following operations and/or methods to the one or more traces: determination of a linear slope, determination of a plurality of non-linear parameters describing curvature of the one or more traces, determination of an abrupt change in intensity of the one or more traces, determination of one or more changes in a baseline intensity of the one or more traces, determination of a change of a frequency-domain representation of the one or more traces, determination of a change of the power-spectral domain representation of the one or more traces, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, or determination of a maximum Lyapunov exponent.

[0009] In some embodiments, the trained model is selected from the group consisting of: a neural network algorithm, a support vector machine algorithm, a decision tree algorithm, an unsupervised clustering algorithm, a supervised clustering algorithm, a regression algorithm, a gradient-boosting algorithm (*e.g.*, a gradient-boosting implementation of a machine learning algorithm such as gradient-boosted decision trees) and any combination thereof. In some embodiments, the trained model comprises a gradient-boosted ensemble model. In some embodiments, the trained model is configured to process one or more features selected from the

group consisting of recurrence rates, determinism, mean diagonal length, maximum diagonal length, divergence, Shannon entropy in diagonal length, trend in recurrences, laminarity, trapping time, maximum vertical line length, Shannon entropy in vertical line lengths, mean recurrence time, Shannon entropy in recurrence times, number of the most probable recurrences, and/or any combination thereof. In some embodiments, the trained model is configured to process two or more features selected from the group consisting of recurrence rates, determinism, mean diagonal length, maximum diagonal length, divergence, Shannon entropy in diagonal length, trend in recurrences, laminarity, trapping time, maximum vertical line length, Shannon entropy in vertical line lengths, mean recurrence time, Shannon entropy in recurrence times, number of the most probable recurrences, and/or any combination thereof.

[0010] In some embodiments, the trained model is configured to process one or more features of the temporal dynamic of one or more traces. In some embodiments, the temporal dynamics of the one or more traces are determined by data analysis methods. In some embodiments, the data analysis methods apply one or more of the following operations and/or methods to the one or more traces: determination of a linear slope, determination of a plurality of non-linear parameters describing curvature of the one or more traces, determination of an abrupt change in intensity of the one or more traces, determination of one or more changes in a baseline intensity of the one or more traces, determination of a change of a frequency-domain representation of the one or more traces, determination of a change of the power-spectral domain representation of the one or more traces, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, or determination of a maximum Lyapunov exponent.

[0011] In some embodiments, the method further comprises predicting a subject's diagnostic status with respect to the disease or disorder using a model that has a sensitivity of at least about 70%, 75%, 80%, 85% or 90% at predicting diagnostic status with respect to the disease or disorder across a suitable cohort population (*e.g.*, such as the one provided in in the Examples section below).

[0012] In some embodiments, the method further comprises predicting a subject's diagnostic status with respect to the disease or disorder using a model that has a sensitivity of up to about

70%, 75%, 80%, 85% or 90% at predicting diagnostic status with respect to the disease or disorder across a suitable cohort population.

[0013] In some embodiments, the method further comprises predicting a subject's diagnostic status with respect to the disease or disorder using a model that has a specificity of at least about 70%, 75%, 80%, 85% or 90% at predicting diagnostic status with respect to the disease or disorder across a suitable cohort population.

[0014] In some embodiments, the method further comprises predicting a subject's diagnostic status with respect to the disease or disorder using a model that has a specificity of up to about 70%, 75%, 80%, 85% or 90% at predicting diagnostic status with respect to the disease or disorder across a suitable cohort population.

[0015] In some embodiments, the method further comprises predicting a subject's diagnostic status with respect to the disease or disorder with a model that has a positive predictive value of at least about 70%, 75%, 80%, 85% or 90% at predicting diagnostic status with respect to the disease or disorder across a suitable cohort population.

[0016] In some embodiments, the method further comprises predicting a subject's diagnostic status with respect to the disease or disorder with a model that has a positive predictive value of up to about 70%, 75%, 80%, 85% or 90% at predicting diagnostic status with respect to the disease or disorder across a suitable cohort population.

[0017] In some embodiments, the method further comprises predicting a subject's diagnostic status with respect to the disease or disorder with a model that has a negative predictive value of at least about 70%, 75%, 80%, 85% or 90% at predicting diagnostic status with respect to the disease or disorder across a suitable cohort population.

[0018] In some embodiments, the method further comprises predicting a subject's diagnostic status with respect to the disease or disorder with a model that has a negative predictive value of up to about 70%, 75%, 80%, 85% or 90% at predicting diagnostic status with respect to the disease or disorder across a suitable cohort population.

[0019] In some embodiments, the method further comprises predicting a subject's diagnostic status with respect to a disease or disorder with a model that predicts diagnostic status with respect to the disease or disorder with an Area Under the Receiver Operating Characteristic (AUROC) of at least about 0.65, at least about 0.70, at least about 0.75, at least about 0.80, at least about 0.82, at least about 0.84, at least about 0.86, at least about 0.88, or at least about 0.90 with respect to a suitable cohort population.

[0020] In another aspect, the present disclosure provides a device comprising one or more processors, and memory storing one or more programs for execution by the one or more processors, the one or more programs comprising instructions for: (a) sampling each respective position in a plurality of positions along a reference line on a biological sample of a subject associated with a Raman signature of the subject, thereby obtaining a plurality of Raman spectra, each Raman spectrum in the plurality of Raman spectra corresponding to a different position in the plurality of positions, and each position in the plurality of positions representing a different period of growth of the biological sample associated with the Raman signature; (b) analyzing each of the plurality of Raman spectra across a reference line on the biological sample thereby obtaining a first dataset; (c) deriving a respective second dataset from the corresponding plurality of the Raman spectra measurements, each respective feature in the corresponding set of features being determined by a sequential variation in the Raman spectra; and (d) processing the features using a trained model to predict a subject's diagnostic status with respect to disease or disorder associated with the Raman signature. In some embodiments, the respective second dataset is derived by applying recurrence quantification analysis or related methods to the corresponding plurality of Raman spectra measurements. In some embodiments, the analyzing of the Raman spectra comprises cosmic ray removal, background correction, normalization, peak fitting, or any combination thereof.

[0021] In some embodiments, the biological sample comprises a tooth sample, a hair sample, a nail sample, or any combination thereof. In some embodiments, the instructions further comprise detecting or monitoring changes in the Raman spectra across the plurality of positions indicative of a temporal response of the subject. In some embodiments, the temporal response comprises a biological response, a physiological response, an anatomical response, a treatment response, a stress-related response, or a combination thereof response. In some embodiments, the plurality of Raman spectra comprises from about 200 to about 3700 wave numbers. In some embodiments, sampling comprises using a Raman spectroscopy microscope. In some embodiments, the Raman spectroscopy microscope comprises an 50X air coupled objective, 63X water immersion coupled objection, or any combination thereof. In some embodiments, sampling comprises exposing the biological sample to a light source to generate the Raman spectra of the plurality of Raman spectra at the plurality of positions. In some embodiments, the light source comprises a laser, wherein the laser comprises a wavelength of about 785 nm, a wavelength of about 532 nm, or any combination thereof. In some embodiments, the instructions further

comprise translating, wherein translating comprises moving the biological sample with a step size of about 2 microns to about 5 microns from a first position to a second position of the plurality of positions subsequent to acquiring a Raman spectrum of the plurality of Raman spectra. In some embodiments, translating is performed using an integration time of about 0.2 seconds to about 0.3 seconds. In some embodiments, the disease or disorder comprises autism spectrum disorder (ASD), attention deficit/hyperactivity disorder (ADHD), amyotrophic lateral sclerosis (ALS), schizophrenia, irritable bowel disease (IBD), pediatric kidney disease, kidney transplant rejection, pediatric cancer or any combination thereof. In some embodiments, the disease or disorder comprises the ASD. In some embodiments, predicting a subject's diagnostic status with respect to a disease or disorder comprises processing changes in the Raman spectra across the plurality of positions with a trained model. In some embodiments, the trained model is selected from the group consisting of: a neural network algorithm, a support vector machine algorithm, a decision tree algorithm, an unsupervised clustering algorithm, a supervised clustering algorithm, a regression algorithm, a gradient-boosting algorithm, and any combination thereof. In some embodiments, the trained model comprises a gradient-boosted ensemble model. In some embodiments, the trained model is configured to process one or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, and any combination thereof. In some embodiments, the trained model is configured to process two or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, and any combination thereof.

[0022] In some embodiments, the trained model is configured to process one or more features of the temporal dynamic of one or more traces. In some embodiments, the temporal dynamics of the one or more traces are determined by data analysis methods. In some embodiments, the data analysis methods apply one or more of the following operations and/or methods to the one or more traces: determination of a linear slope, determination of a plurality of non-linear parameters describing curvature of the one or more traces, determination of an abrupt change in intensity of the one or more traces, determination of one or more changes in a baseline intensity of the one or more traces, determination of a change of a frequency-domain representation of the one or more traces, determination of a change of the power-spectral domain representation of the one or more traces, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more

joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, or determination of a maximum Lyapunov exponent.

[0023] In another aspect, the present disclosure provides a non-transitory computer readable storage medium and one or more computer programs embedded therein for classification, the one or more computer programs comprising instructions which, when executed by a computer system, cause the computer system to perform a method comprising: (a) sampling each respective position in a plurality of positions along a reference line on a biological sample of a subject associated with a Raman signature of the subject, thereby obtaining a plurality of Raman spectra, each Raman spectrum in the plurality of Raman spectra corresponding to a different position in the plurality of positions, and each position in the plurality of positions representing a different period of growth of the biological sample associated with the Raman signature; (b) analyzing each of the plurality of Raman spectra across a reference line on the biological sample thereby obtaining a first dataset; (c) deriving a respective second dataset from the corresponding plurality of the Raman spectra measurements, each respective feature in the corresponding set of features being determined by a sequential variation in the Raman spectra; and (d) processing the features using a trained model to predict a subject's diagnostic status with respect to disease or disorder associated with the Raman signature. In some embodiments, the respective second dataset is derived by applying recurrence quantification analysis or related methods to the corresponding plurality of Raman spectra measurements. In some embodiments, the analyzing of the Raman spectra comprises cosmic ray removal, background correction, normalization, peak fitting, or any combination thereof.

[0024] In some embodiments, the biological sample comprises a tooth sample, a hair sample, a nail sample, or any combination thereof. In some embodiments, the method further comprises detecting or monitoring changes in the Raman spectra across the plurality of positions indicative of a temporal response (i.e., one or more traces) of the subject. In some embodiments, the temporal response comprises a biological response, a physiological response, an anatomical response, a treatment response, a stress-related response, or a combination thereof response. In some embodiments, the plurality of Raman spectra comprises from about 200 to about 3700 wave numbers. In some embodiments, sampling comprises using a Raman spectroscopy microscope. In some embodiments, the Raman spectroscopy microscope comprises an 50X air coupled objective, 63X water immersion coupled objection, or any combination thereof. In some

embodiments, sampling comprises exposing the biological sample to a light source to generate the Raman spectra of the plurality of Raman spectra at the plurality of positions. In some embodiments, the light source comprises a laser, wherein the laser comprises a wavelength of about 785 nm, a wavelength of about 532 nm, or any combination thereof. In some embodiments, the instructions further comprise translating, wherein translating comprises moving the biological sample with a step size of about 2 microns to about 5 microns from a first position to a second position of the plurality of positions subsequent to acquiring a Raman spectrum of the plurality of Raman spectra. In some embodiments, translating is performed using an integration time of about 0.2 seconds to about 0.3 seconds. In some embodiments, the disease or disorder comprises autism spectrum disorder (ASD), attention deficit/hyperactivity disorder (ADHD), amyotrophic lateral sclerosis (ALS), schizophrenia, irritable bowel disease (IBD), pediatric kidney disease, kidney transplant rejection, pediatric cancer or any combination thereof. In some embodiments, the disease or disorder comprises the ASD. In some embodiments, predicting a subject's diagnostic status with respect to the disease or disorder comprises processing changes in the Raman spectra across the plurality of positions with a trained model. In some embodiments, the trained model is selected from the group consisting of: a neural network algorithm, a support vector machine algorithm, a decision tree algorithm, an unsupervised clustering algorithm, a supervised clustering algorithm, a regression algorithm, a gradient-boosting algorithm, and any combination thereof. In some embodiments, the trained model comprises a gradient-boosted ensemble model. In some embodiments, the trained model is configured to process one or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, and any combination thereof. In some embodiments, the trained model is configured to process two or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, and any combination thereof.

[0025] In some embodiments, the trained model is configured to process one or more features of the temporal dynamic of one or more traces. In some embodiments, the temporal dynamics of the one or more traces are determined by data analysis methods. In some embodiments, the data analysis methods apply one or more of the following operations and/or methods to the one or more traces: determination of a linear slope, determination of a plurality of non-linear parameters describing curvature of the one or more traces, determination of an abrupt change in intensity of the one or more traces, determination of one or more changes in a baseline intensity of the one or

more traces, determination of a change of a frequency-domain representation of the one or more traces, determination of a change of the power-spectral domain representation of the one or more traces, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, or determination of a maximum Lyapunov exponent.

[0026] In another aspect, the present disclosure provides a method for training a model, comprising: at a computer system having one or more processors, and memory storing one or more programs for execution by the one or more processors: (a) for each respective training subject in a plurality of training subjects, wherein a first subset of training subjects in the plurality of training subjects have a first diagnostic status corresponding to having a first biological condition associated with a Raman signature and a second subset of training subjects in the plurality of training subjects have a second diagnostic status corresponding to not having the first biological condition associated with the Raman signature: (i) sampling each respective position in a plurality of positions along a reference line on a biological sample of the subject associated with the Raman signature of the subject, thereby obtaining a plurality of Raman spectra, each Raman spectrum in the plurality of Raman spectra corresponding to a different position in the plurality of positions, and each position in the plurality of positions represent a different period of growth of the biological sample of the subject associated with the Raman signature; (ii) analyzing each Raman spectrum across reference line on biological sample thereby obtaining a first dataset; and (iii) deriving a respective second dataset from the corresponding plurality of Raman spectra, each respective feature in the corresponding set of features being determined by a sequential variation in Raman spectra; and (b) training an untrained or partially untrained model with (i) the corresponding set of features of each respective second dataset of each training subject in the plurality of training subjects and (ii) the corresponding diagnostic status of each training subject in the plurality of training subjects, selected from among the first diagnostic status and the second diagnostic status, thereby obtaining a trained model that provides an indication as to whether a test subject has the first biological condition associated with the Raman signature based on values for features in a set of features acquired from a biological sample associated with the Raman signature of the test subject. In some embodiments, the respective second dataset is derived by applying recurrence quantification analysis or related

methods to the corresponding plurality of Raman spectra measurements. In some embodiments, the analyzing of the Raman spectra comprises cosmic ray removal, background correction, normalization, peak fitting, or any combination thereof.

[0027] In some embodiments, the trained model is a neural network algorithm, a support vector machine algorithm, a decision tree algorithm, an unsupervised clustering model algorithm, a supervised clustering model algorithm, a regression model, or a gradient-boosting algorithm (*e.g.*, a gradient-boosting implementation of a machine learning algorithm such as gradient-boosted decision trees). In some embodiments, the trained model is a multinomial classifier. In some embodiments, the trained model is a binomial classifier. In some embodiments, the trained model is a regressor. In some embodiments, the first biological condition is selected from the group consisting of autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD), amyotrophic lateral sclerosis (ALS), schizophrenia, irritable bowel disease (IBD), pediatric kidney disease, kidney transplant rejection, and pediatric cancer.

[0028] In some embodiments, evaluating the test subject for the first biological condition associated with a Raman signature further includes discriminating between a presence of the first biological condition associated with the Raman signature and an absence of the first biological condition associated with the Raman signature. In some embodiments, evaluating the test subject for the first biological condition associated with the Raman signature further includes discriminating between the first biological condition associated with the Raman signature and a second biological condition associated with the Raman signature distinct from the first biological condition associated with the Raman signature. In some embodiments, the first biological condition is autism spectrum disorder and the second biological condition is neurotypical development; that is, the absence of a neurodevelopmental disorder. In some embodiments, the first biological condition is autism spectrum disorder and the second biological condition is attention-deficit/hyperactivity disorder. In some embodiments, the test subject is a human. In some embodiments, the test subject is an adult. In some embodiments, the human is between the ages of about 12 and about 5 years old. In some embodiments, the subject is less than about 12, 11, 10, 9, 8, 7, 5, 4, 3, 2, or 1 year(s) old. In some embodiments, the subject is at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 year(s) old. In some embodiments, at least a portion of the temporal profile of the Raman profile corresponds to a prenatal period of the subject.

[0029] In some embodiments, the corresponding biological sample associated with the Raman signature of the respective training subject is selected from the group consisting of a hair

shaft, a tooth, and a nail. In some embodiments, the corresponding biological sample associated with the Raman signature of the respective training subject is the hair shaft, and wherein the reference line corresponds to a longitudinal direction of the hair shaft. In some embodiments, the corresponding biological sample associated with the Raman signature of the respective training subject is the tooth, and wherein the reference line corresponds to a direction across the growth bands, including the neonatal line of the tooth. In some embodiments, the corresponding plurality of positions is sequenced such that a first position in the corresponding plurality of positions along the corresponding biological sample of the respective training subject corresponds to a position closest to a tip of the corresponding biological sample of the respective training subject. In some embodiments, each trace in the corresponding plurality of Raman spectra measurements includes a plurality of data points, each data point being an instance of the respective position in the plurality of positions. In some embodiments, the corresponding set of features is selected from the group consisting of recurrence rates, determinism, mean diagonal length, maximum diagonal length, divergence, Shannon entropy in diagonal length, trend in recurrences, laminarity, trapping time, maximum vertical line length, Shannon entropy in vertical line lengths, mean recurrence time, Shannon entropy in recurrence times, number of the most probable recurrences, and/or any combination thereof. In some embodiments, the corresponding plurality of positions includes at least 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, 10000, or more than 10000 positions.

[0030] In some embodiments, the corresponding set of features are selected from a group of temporal dynamic features of one or more traces. In some embodiments, the temporal dynamic features of the one or more traces are determined by data analysis methods. In some embodiments, the data analysis methods apply one or more of the following operations and/or methods to one or more traces: determination of a linear slope, determination of a plurality of non-linear parameters describing curvature of the one or more traces, determination of an abrupt change in intensity of the one or more traces, determination of one or more changes in a baseline intensity of the one or more traces, determination of a change of a frequency-domain representation of the one or more traces, determination of a change of the power-spectral domain representation of the one or more traces, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters,

determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, or determination of a maximum Lyapunov exponent.

[0031] Another aspect of the present disclosure provides a non-transitory computer readable medium comprising machine executable code that, upon execution by one or more computer processors, implements any of the methods above or elsewhere herein.

[0032] Another aspect of the present disclosure provides a system comprising one or more computer processors and computer memory coupled thereto. The computer memory comprises machine executable code that, upon execution by the one or more computer processors, implements any of the methods above or elsewhere herein.

[0033] Additional aspects and advantages of the present disclosure will become readily apparent to those skilled in this art from the following detailed description, wherein only illustrative embodiments of the present disclosure are shown and described. As will be realized, the present disclosure is capable of other and different embodiments, and its several details are capable of modifications in various obvious respects, all without departing from the disclosure. Accordingly, the drawings and description are to be regarded as illustrative in nature, and not as restrictive.

INCORPORATION BY REFERENCE

[0034] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference. To the extent publications and patents or patent applications incorporated by reference contradict the disclosure contained in the specification, the specification is intended to supersede and/or take precedence over any such contradictory material.

BRIEF DESCRIPTION OF THE DRAWINGS

[0035] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings (also “Figure” and “FIG.” herein), of which:

[0036] **FIG. 1** shows an example of a block diagram of a computing device 100 of the present disclosure.

[0037] FIGS. 2A-2C show illustrations of a hair sample (FIG. 2A), a tooth sample (FIG. 2B), and a nail sample (FIG. 2C) of a subject.

[0038] FIG. 3 shows a flow chart of a method 300 for evaluating a subject for a biological condition.

[0039] FIG. 4 shows a computer system that is programmed or otherwise configured to implement methods provided herein.

[0040] FIG. 5 shows an example of model accuracy for predicting diagnostic status for autism spectrum disorder (ASD) utilizing features derived from application of RQA to ICA-derived dimensions of the Raman waveform, as indicated by an experimental Receiver Operating Characteristics (ROC) curve for evaluating accuracy of the disclosed method of evaluating a subject for autism spectrum disorder. Device performance is measured by calculating the area-under-the-curve (AUC) of the ROC plot, which provides a measure of performance at varying classification thresholds; here, the AUC was 0.86, indicating robustly accurate predictive performance.

[0041] FIG. 6 shows an example of model accuracy for predicting diagnostic status for amyotrophic lateral sclerosis (ALS) utilizing features derived from application of RQA to ICA-derived dimensions of the Raman waveform, as indicated by an experimental Receiver Operating Characteristics (ROC) curve for evaluating accuracy of the disclosed method of evaluating a subject for autism spectrum disorder. Device performance is measured by calculating the area-under-the-curve (AUC) of the ROC plot, which provides a measure of performance at varying classification thresholds; here, the AUC was 0.88, indicating robustly accurate predictive performance.

DETAILED DESCRIPTION

[0042] While various embodiments of the invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions may occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed.

[0043] Dynamic biological responses may be indicative of underlying biological processes having structural and functional significance for humans. For example, aberrant or abnormal

dynamic biological response may be associated with many biological conditions, such as diseases and disorders. Examples of such biological conditions may include neurological conditions (*e.g.*, autism spectrum disorder, schizophrenia, or attention-deficit/hyperactivity disorder (ADHD)), neurodegenerative conditions (*e.g.*, amyotrophic lateral sclerosis (ALS), Alzheimer's disease, Parkinson's disease, and Huntington's disease), and cancers (*e.g.*, pediatric cancer).

[0044] Given the above background, there is a need for accurate methods and systems for the diagnosis of biological conditions, and especially for non-invasive diagnosis. Such diagnosis may be based on accurate profiling of biomarkers detectable with non-invasive methods for diagnosis of the biological conditions. The present disclosure provides improved systems and methods for accurate diagnosis of biological conditions based on analysis of dynamic biological response data from non-invasively obtained biological samples from subjects. Such improved systems and methods for accurate diagnosis of biological conditions, in some embodiments, are based on a combination of Raman profiling of biological samples and artificial intelligence data analysis. The present disclosure addresses these needs, for example, by providing a biological sample biomarker for diagnosis of biological conditions. The biological sample includes a human biological specimen that is associated with incremental growth. Such a biological sample could be a hair shaft, a tooth, and a nail. The non-invasive biomarker of the present disclosure can be used for the diagnosis of young children, even infants younger than one year old. In some cases, the child is between the ages of about 12 and about 5 years old. In some embodiments, the child is less than about 12, 11, 10, 9, 8, 7, 5, 4, 3, 2, or 1 year(s) old. In some embodiments, the child is at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 year(s) old.

[0045] In an aspect, the present disclosure provides a method for predicting a subject's diagnostic status with respect to a disease or disorder, comprising: (a) exposing a biological sample of the subject to a light source, where the biological sample comprises a tooth sample, a hair sample, or a nail sample; (b) acquiring a plurality of Raman spectra from the exposed biological sample; (c) processing the plurality of Raman spectra to generate a spatial map of the plurality of Raman spectra; and (d) predicting a subject's diagnostic status with respect to a disease or disorder based at least in part on the spatial map of the plurality of Raman spectra. In some embodiments, the light source comprises a laser.

[0046] In some embodiments, the analyzing determines temporal dynamics of underlying biological processes. In some embodiments, the analyzing comprises reducing the dimensionality of the plurality of Raman spectra (*e.g.*, by independent components analysis) prior to the

processing. In some embodiments, the optical signal is generated by a light source (*e.g.*, a laser). In some embodiments, the biological sample comprises the tooth sample. In some embodiments, the method further comprises detecting or monitoring changes in a temporal stress profile (*i.e.*, one or more traces described elsewhere herein) that are indicative of a temporal response of the subject. In some embodiments, the temporal response comprises a biochemical response. In some embodiments, the temporal response comprises a biological response, a physiological response, an anatomical response, a treatment response, a stress-related response, or a combination thereof. In some embodiments, the plurality of Raman spectra comprises from about 200 to about 3700 wave numbers. In some embodiments, the acquiring comprises using Raman spectroscopy microscope. In some embodiments, the Raman spectroscopy microscope comprises an 50X air coupled objective, 63X water immersion coupled objection, or any combination thereof. In some embodiments, the laser comprises a wavelength of about 785 nm, a wavelength of about 532 nm, or any combination thereof. In some embodiments, the acquiring is performed using an integration time of about 0.2 seconds to about 0.3 seconds. In some embodiments, the acquiring comprises moving the biological sample with a step size of about 2 microns to about 5 microns, subsequent to acquiring a Raman spectrum of the plurality of Raman spectra.

[0047] In some embodiments, the systems and methods disclosed herein use Raman Spectroscopy alone, or in combination with other techniques. Such techniques, in some embodiments, include laser ablation-inductively coupled plasma-mass spectrometry (LA-ICP-MS), C-reactive immunohistochemistry fluorescence staining, and others. In some embodiments, combining techniques improves diagnostic accuracy or precision of a given technique alone. In some embodiments, the addition of LA-ICP-MS provides a plurality of non-invasive metal metabolism biomarkers of a given biological sample that complement the diagnostic power of Raman Spectroscopy. In some embodiments, the metal metabolism biomarkers comprise Zinc, Tin, Magnesium, Copper, Iodide, lithium, aluminum, phosphorus, sulfur, calcium, chromium, manganese, iron, cobalt, nickel, arsenic, strontium, cadmium, tin, iodine, barium, mercury, lead, bismuth, molybdenum, or any combination thereof. In some embodiments, the addition of C-reactive protein immunohistochemistry fluorescence provides temporal fluctuations of inflammation to complement the diagnostic power of Raman Spectroscopy. In some embodiments, the plurality of metal metabolism biomarkers includes at least 2, at least 5, or at least 10 metal metabolism biomarkers. In some embodiments, the plurality of metal metabolism biomarkers includes no more than 20, no more than 10, or no more than 5 metal metabolism

biomarkers. In some embodiments, the plurality of metal metabolism biomarkers consists of from 2 to 5, from 3 to 10, or from 8 to 20 metal metabolism biomarkers. In some embodiments, the plurality of metal metabolism biomarkers falls within another range starting no lower than 2 and ending no higher than 20 metal metabolism biomarkers. In some embodiments, the plurality of Raman spectra includes at least 2, at least 5, or at least 10 Raman spectra. In some embodiments, the plurality of Raman spectra includes no more than 20, no more than 10, or no more than 5 Raman spectra. In some embodiments, the plurality of Raman spectra consists of from 2 to 5, from 3 to 10, or from 8 to 20 Raman spectra. In some embodiments, the plurality of Raman spectra falls within another range starting no lower than 2 and ending no higher than 20 Raman spectra.

[0048] In some embodiments, the disease or disorder comprises autism spectrum disorder (ASD), attention deficit/hyperactivity disorder (ADHD), amyotrophic lateral sclerosis (ALS), schizophrenia, irritable bowel disease (IBD), pediatric kidney disease, kidney transplant rejection, pediatric cancer or any combination thereof. In some embodiments, the disease or disorder comprises the ASD. In some embodiments, the subject is a human. In some embodiments, the subject is an adult. In some embodiments, the subject is between the ages of about 12 and about 5 years old. In some embodiments, the subject is less than about 12, 11, 10, 9, 8, 7, 5, 4, 3, 2, or 1 year(s) old. In some embodiments, the subject is at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 year(s) old. In some embodiments, at least a portion of the temporal Raman profile corresponds to a prenatal period of the subject.

[0049] In some embodiments, predicting a subject's diagnostic status with respect to a disease or disorder comprises processing the spatial map using a trained model. In some embodiments, the processing comprises extracting features from the spatial map (*e.g.*, by recurrence quantification analysis), and analyzing the features using the trained model. In some embodiments, the processing comprises computational analysis of temporal dynamics derived from the spatial map, *e.g.*, by application of dimensionality reduction techniques, including independent component analysis (ICA) and/or principal component analysis (PCA), followed by the subsequent application of recurrence quantification analysis (RQA) to extract computational features descriptive of the dimensions derived from ICA/PCA.

[0050] In some embodiments, the trained model comprises a plurality of parameters, where the term "parameter" refers to any coefficient or, similarly, any value of an internal or external element (*e.g.*, a weight and/or a hyperparameter) in the model (*e.g.*, where the model is a

regressor or a classifier) that can affect (*e.g.*, modify, tailor, and/or adjust) one or more inputs, outputs, and/or functions in the model. For example, in some embodiments, a parameter of a model refers to any coefficient, weight, and/or hyperparameter that can be used to control, modify, tailor, and/or adjust the behavior, learning, and/or performance of the model. In some instances, a parameter is used to increase or decrease the influence of an input (*e.g.*, a feature) to a model. As a nonlimiting example, in some embodiments, a parameter is used to increase or decrease the influence of a node (*e.g.*, of a neural network), where the node includes one or more activation functions. Assignment of parameters to specific inputs, outputs, and/or functions of a model is not limited to any one paradigm for a given model but can be used in any suitable model for a desired performance. In some embodiments, a parameter has a fixed value. In some embodiments, a value of a parameter is manually and/or automatically adjustable. In some embodiments, a value of a parameter is modified by a validation and/or training process for a model (*e.g.*, by error minimization and/or back propagation methods). In some embodiments, a model of the present disclosure includes a plurality of parameters. In some embodiments, the plurality of parameters associated with a model (*e.g.*, an untrained, partially trained, or fully trained model) is n parameters, where: $n \geq 2$; $n \geq 5$; $n \geq 10$; $n \geq 25$; $n \geq 40$; $n \geq 50$; $n \geq 75$; $n \geq 100$; $n \geq 125$; $n \geq 150$; $n \geq 200$; $n \geq 225$; $n \geq 250$; $n \geq 350$; $n \geq 500$; $n \geq 600$; $n \geq 750$; $n \geq 1,000$; $n \geq 2,000$; $n \geq 4,000$; $n \geq 5,000$; $n \geq 7,500$; $n \geq 10,000$; $n \geq 20,000$; $n \geq 40,000$; $n \geq 75,000$; $n \geq 100,000$; $n \geq 200,000$; $n \geq 500,000$, $n \geq 1 \times 10^6$, $n \geq 5 \times 10^6$, or $n \geq 1 \times 10^7$. In some embodiments n is between 10,000 and 1×10^7 , between 100,000 and 5×10^6 , or between 500,000 and 1×10^6 . In some embodiments, the plurality of parameters includes at least 10, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1×10^6 , at least 5×10^6 , at least 1×10^7 , at least 5×10^7 , or at least 1×10^8 parameters. In some embodiments, the plurality of parameters includes no more than 1×10^9 , no more than 1×10^8 , no more than 1×10^7 , no more than 1×10^6 , no more than 100,000, no more than 10,000, no more than 1000, or no more than 100 parameters. In some embodiments, the plurality of parameters consists of from 10 to 10,000, from 100 to 100,000, from 1000 to 1×10^6 , from 100,000 to 1×10^7 , from 1×10^6 to 1×10^8 , or from 1×10^7 to 1×10^9 parameters. In some embodiments, the plurality of parameters falls within another range starting no lower than 10 and ending no higher than 1×10^9 parameters.

[0051] In some embodiments, the trained model is selected from the group consisting of: a neural network algorithm, a support vector machine algorithm, a decision tree algorithm, an unsupervised clustering algorithm, a supervised clustering algorithm, a regression algorithm, a

gradient-boosting algorithm (*e.g.*, a gradient-boosting implementation of a machine learning algorithm such as gradient-boosted decision trees) and any combination thereof. In some embodiments, the trained model comprises a gradient-boosted ensemble model. In some embodiments, the trained model is configured to process one or more features selected from the group consisting of recurrence rates, determinism, mean diagonal length, maximum diagonal length, divergence, Shannon entropy in diagonal length, trend in recurrences, laminarity, trapping time, maximum vertical line length, Shannon entropy in vertical line lengths, mean recurrence time, Shannon entropy in recurrence times, number of the most probable recurrences, and/or any combination thereof. In some embodiments, the trained model is configured to process two or more features selected from the group consisting of recurrence rates, determinism, mean diagonal length, maximum diagonal length, divergence, Shannon entropy in diagonal length, trend in recurrences, laminarity, trapping time, maximum vertical line length, Shannon entropy in vertical line lengths, mean recurrence time, Shannon entropy in recurrence times, number of the most probable recurrences, and/or any combination thereof.

[0052] In some embodiments, the trained model is configured to process one or more features of the temporal dynamic of one or more traces. In some embodiments, the temporal dynamics of the one or more traces are determined by data analysis methods. In some embodiments, the data analysis methods apply one or more of the following operations and/or methods to the one or more traces: determination of a linear slope, determination of a plurality of non-linear parameters describing curvature of the one or more traces, determination of an abrupt change in intensity of the one or more traces, determination of one or more changes in a baseline intensity of the one or more traces, determination of a change of a frequency-domain representation of the one or more traces, determination of a change of the power-spectral domain representation of the one or more traces, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, or determination of a maximum Lyapunov exponent.

[0053] In some embodiments, the method further comprises predicting a subject's diagnostic status with respect to the disease or disorder with a sensitivity of at least about 80%. In some embodiments, the method further comprises predicting a subject's diagnostic status with respect to the disease or disorder with a sensitivity of up to about 80%. In some embodiments, the

method further comprises predicting a subject's diagnostic status with respect to the disease or disorder with a specificity of at least about 80%. In some embodiments, the method further comprises predicting a subject's diagnostic status with respect to the disease or disorder with a specificity of up to about 80%. In some embodiments, the method further comprises predicting a subject's diagnostic status with respect to the disease or disorder with a positive predictive value of at least about 80%. In some embodiments, the method further comprises predicting a subject's diagnostic status with respect to the disease or disorder with a positive predictive value of up to about 80%. In some embodiments, the method further comprises predicting a subject's diagnostic status with respect to the disease or disorder with a negative predictive value of at least about 80%. In some embodiments, the method further comprises predicting a subject's diagnostic status with respect to the disease or disorder with a negative predictive value of up to about 80%. In some embodiments, the method further comprises predicting a subject's diagnostic status with respect to the disease or disorder with an Area Under the Receiver Operating Characteristic (AUROC) of at least about 0.80.

[0054] In another aspect, the present disclosure provides a device comprising one or more processors, and memory storing one or more programs for execution by the one or more processors, the one or more programs comprising instructions for: (a) sampling each respective position in a plurality of positions along a reference line on a biological sample of a subject associated with a Raman signature of the subject, thereby obtaining a plurality of Raman spectra, each Raman spectrum in the plurality of Raman spectra corresponding to a different position in the plurality of positions, and each position in the plurality of positions representing a different period of growth of the biological sample associated with the Raman signature; (b) analyzing each of the plurality of Raman spectra across a reference line on the biological sample thereby obtaining a first dataset; (c) deriving a respective second dataset from the corresponding plurality of the Raman spectra measurements, each respective feature in the corresponding set of features being determined by a sequential variation in the Raman spectra; and (d) processing the features using a trained model to predict a subject's diagnostic status with respect to disease or disorder associated with the Raman signature. In some embodiments, the respective second dataset is derived by applying recurrence quantification analysis or related methods to the corresponding plurality of Raman spectra measurements. In some embodiments, the analyzing of the Raman spectra comprises cosmic ray removal, background correction, normalization, peak fitting, or any combination thereof.

[0055] In another aspect, the present disclosure provides a non-transitory computer readable storage medium and one or more computer programs embedded therein for classification, the one or more computer programs comprising instructions which, when executed by a computer system, cause the computer system to perform a method comprising: (a) sampling each respective position in a plurality of positions along a reference line on a biological sample of a subject associated with a Raman signature of the subject, thereby obtaining a plurality of Raman spectra, each Raman spectrum in the plurality of Raman spectra corresponding to a different position in the plurality of positions, and each position in the plurality of positions representing a different period of growth of the biological sample associated with the Raman signature; (b) analyzing each of the plurality of Raman spectra across a reference line on the biological sample thereby obtaining a first dataset; (c) deriving a respective second dataset from the corresponding plurality of the Raman spectra measurements, each respective feature in the corresponding set of features being determined by a sequential variation in the Raman spectra; and (d) processing the features using a trained model to predict a subject's diagnostic status with respect to disease or disorder associated with the Raman signature. In some embodiments, the respective second dataset is derived by applying recurrence quantification analysis or related methods to the corresponding plurality of Raman spectra measurements. In some embodiments, the analyzing of the Raman spectra comprises cosmic ray removal, background correction, normalization, peak fitting, or any combination thereof.

[0056] In another aspect, the present disclosure provides a method for training a model, comprising: at a computer system having one or more processors, and memory storing one or more programs for execution by the one or more processors: (a) for each respective training subject in a plurality of training subjects, wherein a first subset of training subjects in the plurality of training subjects have a first diagnostic status corresponding to having a first biological condition associated with a Raman signature and a second subset of training subjects in the plurality of training subjects have a second diagnostic status corresponding to not having the first biological condition associated with the Raman signature: (i) sampling each respective position in a plurality of positions along a reference line on a biological sample of the subject associated with the Raman signature of the subject, thereby obtaining a plurality of Raman spectra, each Raman spectrum in the plurality of Raman spectra corresponding to a different position in the plurality of positions, and each position in the plurality of positions represent a different period of growth of the biological sample of the subject associated with the Raman

signature; (ii) analyzing each Raman spectrum across reference line on biological sample thereby obtaining a first dataset; and (iii) deriving a respective second dataset from the corresponding plurality of Raman spectra, each respective feature in the corresponding set of features being determined by a sequential variation in Raman spectra; and (b) training an untrained or partially untrained model with (i) the corresponding set of features of each respective second dataset of each training subject in the plurality of training subjects and (ii) the corresponding diagnostic status of each training subject in the plurality of training subjects, selected from among the first diagnostic status and the second diagnostic status, thereby obtaining a trained model that provides an indication as to whether a test subject has the first biological condition associated with the Raman signature based on values for features in a set of features acquired from a biological sample associated with the Raman signature of the test subject. In some embodiments, the respective second dataset is derived by applying recurrence quantification analysis or related methods to the corresponding plurality of Raman spectra measurements. In some embodiments, the analyzing of the Raman spectra comprises cosmic ray removal, background correction, normalization, peak fitting, or any combination thereof.

[0057] In some embodiments, a respective subject (*e.g.*, a test subject) is selected from a plurality of subjects. In some embodiments, the plurality of subjects includes at least 2, at least 5, at least 10, at least 20, at least 50, at least 100, or at least 500 subjects. In some embodiments, the plurality of subjects includes no more than 1000, no more than 500, no more than 100, no more than 50, no more than 20, or no more than 10 subjects. In some embodiments, the plurality of subjects consists of from 2 to 10, from 5 to 20, from 10 to 100, or from 100 to 1000 subjects. In some embodiments, the plurality of subjects falls within another range starting no lower than 2 subjects and ending no higher than 1000 subjects.

[0058] In some embodiments, the plurality of training subjects includes at least 2, at least 5, at least 10, at least 20, at least 50, at least 100, at least 500, at least 1000, at least 5000, or at least 100,000 training subjects. In some embodiments, the plurality of training subjects includes no more than 1,000,000, no more than 100,000, no more than 10,000, no more than 1000, no more than 500, no more than 100, no more than 50, no more than 20, or no more than 10 training subjects. In some embodiments, the plurality of training subjects consists of from 2 to 1000, from 500 to 10,000, from 10,000 to 100,000, or from 100,000 to 1,000,000 training subjects. In some embodiments, the plurality of training subjects falls within another range starting no lower than 2 training subjects and ending no higher than 1,000,000 training subjects.

[0059] In some embodiments, a respective subset of training subjects (*e.g.*, the first subset and/or the second subset) in the plurality of training subjects includes at least 2, at least 5, at least 10, at least 20, at least 50, at least 100, at least 500, at least 1000, at least 5000, or at least 10,000 training subjects. In some embodiments, the respective subset of training subjects includes no more than 500,000, no more than 10,000, no more than 1000, no more than 500, no more than 100, no more than 50, no more than 20, or no more than 10 training subjects. In some embodiments, the respective subset of training subjects consists of from 2 to 100, from 50 to 2000, from 1000 to 10,000, or from 10,000 to 500,000 training subjects. In some embodiments, the respective subset of training subjects falls within another range starting no lower than 2 training subjects and ending no higher than 500,000 training subjects.

[0060] In some embodiments, a set of features (*e.g.*, determined by a sequential variation in the Raman spectra) includes at least 1, at least 2, at least 3, at least 5, at least 8, at least 10, at least 15, or at least 20 features. In some embodiments, a set of features includes no more than 50, no more than 20, no more than 10, no more than 5, or no more than 3 features. In some embodiments, a set of features consists of from 1 to 10, from 4 to 15, from 8 to 20, or from 15 to 50 features. In some embodiments, a set of features falls within another range starting no lower than 1 feature and ending no higher than 50 features.

[0061] In some embodiments, the trained model is a neural network algorithm, a support vector machine algorithm, a decision tree algorithm, an unsupervised clustering model algorithm, a supervised clustering model algorithm, a regression model, or a gradient-boosting algorithm (*e.g.*, a gradient-boosting implementation of a machine learning algorithm such as gradient-boosted decision trees). In some embodiments, the trained model is a multinomial classifier. In some embodiments, the trained model is a binomial classifier. In some embodiments, the first biological condition is selected from the group consisting of autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD), amyotrophic lateral sclerosis (ALS), schizophrenia, irritable bowel disease (IBD), pediatric kidney disease, kidney transplant rejection, and pediatric cancer.

[0062] In some embodiments, evaluating the test subject for the first biological condition associated with a Raman signature further includes discriminating between a presence of the first biological condition associated with the Raman signature and an absence of the first biological condition associated with the Raman signature. In some embodiments, evaluating the test subject for the first biological condition associated with the Raman signature further includes

discriminating between the first biological condition associated with the Raman signature and a second biological condition associated with the Raman signature distinct from the first biological condition associated with the Raman signature. In some embodiments, the first biological condition is autism spectrum disorder and the second biological condition is neurotypical development; that is, the absence of a neurodevelopmental disorder. In some embodiments, the first biological condition is autism spectrum disorder and the second biological condition is attention-deficit/hyperactivity disorder. In some embodiments, the test subject is a human. In some embodiments, the test subject is an adult. In some embodiments, the human is between the ages of about 12 and about 5 years old. In some embodiments, the subject is less than about 12, 11, 10, 9, 8, 7, 5, 4, 3, 2, or 1 year(s) old. In some embodiments, the subject is at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 year(s) old. In some embodiments, at least a portion of the temporal profile of the Raman profile corresponds to a prenatal period of the subject. In some embodiments, the human is at least 13, at least 14, at least 15, at least 18, or at least 30 years old. In some embodiments, the human is no more than 40, no more than 30, no more than 18, or no more than 15 years old. In some embodiments, the human is between the ages of about 1 and about 6, between about 5 and about 10, or between about 10 and about 40 years old. In some embodiments, the human falls within another age range starting no lower than about 1 year old and ending no higher than about 40 years old.

[0063] In some embodiments, a first biological condition and/or a second biological condition is selected from a plurality of biological conditions. In some embodiments, the plurality of biological conditions includes at least 2, at least 5, or at least 10 biological conditions. In some embodiments, the plurality of biological conditions includes no more than 20, no more than 10, or no more than 5 biological conditions. In some embodiments, the plurality of biological conditions consists of from 2 to 5, from 3 to 10, or from 8 to 20 biological conditions. In some embodiments, the plurality of biological conditions falls within another range starting no lower than 2 biological conditions and ending no higher than 20 biological conditions. In some embodiments, a first diagnostic status and/or a second diagnostic status is selected from a plurality of diagnostic statuses. In some embodiments, the plurality of diagnostic statuses includes at least 2, at least 5, or at least 10 diagnostic statuses. In some embodiments, the plurality of diagnostic statuses includes no more than 20, no more than 10, or no more than 5 diagnostic statuses. In some embodiments, the plurality of diagnostic statuses consists of from 2 to 5, from 3 to 10, or from 8 to 20 diagnostic statuses. In some embodiments, the plurality of

diagnostic statuses falls within another range starting no lower than 2 diagnostic statuses and ending no higher than 20 diagnostic statuses.

[0064] In some embodiments, the corresponding biological sample associated with the Raman signature of the respective subject (*e.g.*, a test subject and/or a training subject) is selected from the group consisting of a hair shaft, a tooth, and a nail. In some embodiments, the corresponding biological sample associated with the Raman signature of the respective subject (*e.g.*, a test subject and/or a training subject) is the hair shaft, and wherein the reference line corresponds to a longitudinal direction of the hair shaft. In some embodiments, the corresponding biological sample associated with the Raman signature of the respective subject (*e.g.*, a test subject and/or a training subject) is the tooth, and wherein the reference line corresponds to a direction across the growth bands, including the neonatal line of the tooth. In some embodiments, the corresponding plurality of positions is sequenced such that a first position in the corresponding plurality of positions along the corresponding biological sample of the respective subject (*e.g.*, a test subject and/or a training subject) corresponds to a position closest to a tip of the corresponding biological sample of the respective training subject. In some embodiments, each trace in the corresponding plurality of Raman spectra measurements includes a plurality of data points, each data point being an instance of the respective position in the plurality of positions. In some embodiments, the corresponding set of features is selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, and any combination thereof. In some embodiments, the corresponding plurality of positions includes at least 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, 10000, or more than 10000 positions.

[0065] In some embodiments, the plurality of positions along a reference line on a biological sample includes at least 50, at least 100, at least 500, at least 1000, at least 2000, at least 5000, at least 10,000, at least 50,000, at least 100,000, at least 500,000, or at least 1×10^6 positions. In some embodiments, the plurality of positions includes no more than 1×10^7 , no more than 1×10^6 , no more than 100,000, no more than 10,000, no more than 1000, or no more than 100 positions. In some embodiments, the plurality of positions consists of from 50 to 1000, from 500 to 50,000, from 10,000 to 1×10^6 , or from 1×10^6 to 1×10^7 positions. In some embodiments, the plurality of positions falls within another range starting no lower than 50 positions and ending no higher than 1×10^7 positions. In some embodiments, each respective period of growth of the

biological sample is selected from a plurality of periods of growth. In some embodiments, the plurality of periods of growth includes at least 50, at least 100, at least 500, at least 1000, at least 2000, at least 5000, at least 10,000, at least 50,000, at least 100,000, at least 500,000, or at least 1×10^6 periods of growth. In some embodiments, the plurality of periods of growth includes no more than 1×10^7 , no more than 1×10^6 , no more than 100,000, no more than 10,000, no more than 1000, or no more than 100 periods of growth. In some embodiments, the plurality of periods of growth consists of from 50 to 1000, from 500 to 50,000, from 10,000 to 1×10^6 , or from 1×10^6 to 1×10^7 periods of growth. In some embodiments, the plurality of periods of growth falls within another range starting no lower than 50 periods of growth and ending no higher than 1×10^7 periods of growth.

[0066] In some embodiments, the plurality of Raman spectra measurements includes at least 50, at least 100, at least 500, at least 1000, at least 2000, at least 5000, at least 10,000, at least 50,000, at least 100,000, at least 500,000, or at least 1×10^6 Raman spectra measurements. In some embodiments, the plurality of Raman spectra measurements includes no more than 1×10^7 , no more than 1×10^6 , no more than 100,000, no more than 10,000, no more than 1000, or no more than 100 Raman spectra measurements. In some embodiments, the plurality of Raman spectra measurements consists of from 50 to 1000, from 500 to 50,000, from 10,000 to 1×10^6 , or from 1×10^6 to 1×10^7 Raman spectra measurements. In some embodiments, the plurality of Raman spectra measurements falls within another range starting no lower than 50 Raman spectra measurements and ending no higher than 1×10^7 Raman spectra measurements. In some embodiments, a respective Raman spectra measurement includes one or more traces of fluorescent intensity. In some embodiments, a respective Raman spectra measurement includes at least 1, at least 2, at least 3, at least 4, at least 5, or at least 10 traces. In some embodiments, a respective Raman spectra measurement includes no more than 50, no more than 10, no more than 5, or no more than 3 traces. In some embodiments, a respective Raman spectra measurement consists of from 1 to 5, from 2 to 10, or from 10 to 20 traces. In some embodiments, a respective Raman spectra measurement includes another range of traces starting no lower than 1 trace and ending no higher than 20 traces.

[0067] In some embodiments, for a respective trace, the plurality of data points includes at least 2, at least 5, at least 10, at least 20, at least 50, at least 100, at least 500, at least 1000, at least 5000, or at least 10,000 data points. In some embodiments, the plurality of data points includes no more than 500,000, no more than 10,000, no more than 1000, no more than 500, no

more than 100, no more than 50, no more than 20, or no more than 10 data points. In some embodiments, the plurality of data points consists of from 2 to 100, from 50 to 2000, from 1000 to 10,000, or from 10,000 to 500,000 data points. In some embodiments, the plurality of data points falls within another range starting no lower than 2 data points and ending no higher than 500,000 data points.

[0068] In some embodiments, the corresponding set of features are selected from a group of temporal dynamic features of one or more traces. In some embodiments, the temporal dynamic features of the one or more traces are determined by data analysis methods. In some embodiments, the data analysis methods apply one or more of the following operations and/or methods to one or more traces: determination of a linear slope, determination of a plurality of non-linear parameters describing curvature of the one or more traces, determination of an abrupt change in intensity of the one or more traces, determination of one or more changes in a baseline intensity of the one or more traces, determination of a change of a frequency-domain representation of the one or more traces, determination of a change of the power-spectral domain representation of the one or more traces, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, or determination of a maximum Lyapunov exponent.

[0069] Details of an exemplary system are described in conjunction with **FIG. 1**, which shows an example of a block diagram of a computing device 100 of the present disclosure. The device 100 in some implementations includes one or more processing units CPU(s) 102 (also referred to as processors), one or more network interfaces 104, a user interface 106, a non-persistent memory 111, a persistent memory 112, and one or more communication buses 114 for interconnecting these components. The one or more communication buses 114 optionally include circuitry (sometimes called a chipset) that interconnects and controls communications between system components. The non-persistent memory 111 typically includes high-speed random access memory, such as DRAM, SRAM, DDR RAM, ROM, EEPROM, flash memory, whereas the persistent memory 112 typically includes CD-ROM, digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, magnetic disk storage devices, optical disk storage devices, flash memory devices, or other non-volatile solid state storage devices. The persistent memory 112 optionally

includes one or more storage devices remotely located from the CPU(s) 102. The persistent memory 112, and the non-volatile memory device(s) within the non-persistent memory 112, comprise non-transitory computer readable storage medium. In some implementations, the non-persistent memory 111 or alternatively the non-transitory computer readable storage medium stores the following programs, modules and data structures, or a subset thereof, sometimes in conjunction with the persistent memory 112: an optional operating system 116, which includes procedures for handling various basic system services and for performing hardware dependent tasks; an optional network communication module (or instructions) 118 for connecting the system 100 with other devices and/or a communication network 104; an optional classifier training module 120 for training models (*e.g.*, classifiers, regressors, *etc.*) for evaluating a subject for a biological condition; an optional data store 122 for datasets for biological samples from training subjects, including feature data for one or more training subjects 124, where the feature data includes a parameter associated with each of features 126, and diagnostic status 128 (*e.g.*, an indication that a respective training subject has been diagnosed with a biological condition or has not been diagnosed with a biological condition); an optional classifier validation module 130 for validating models that distinguish the a biological condition; an optional data store 132 for datasets for biological samples from validation subjects; and an optional patient classification module 134 for classifying a subject as having a biological condition, *e.g.*, as trained using classifier training module 120.

[0070] In various implementations, one or more of the above identified elements are stored in one or more of the previously mentioned memory devices, and correspond to a set of instructions for performing a function described above. The above identified modules, data, or programs (*e.g.*, sets of instructions) need not be implemented as separate software programs, procedures, datasets, or modules, and thus various subsets of these modules and data may be combined or otherwise re-arranged in various implementations. In some implementations, the non-persistent memory 111 optionally stores a subset of the modules and data structures identified above. Furthermore, in some embodiments, the memory stores additional modules and data structures not described above. In some embodiments, one or more of the above identified elements is stored in a computer system, other than that of visualization system 100, that is addressable by visualization system 100 so that visualization system 100 may retrieve all or a portion of such data when needed.

[0071] In some embodiments, the system 100 is connected to, or includes, one or more analytical devices for performing chemical analyzes. For example, the optional network communication module (or instructions) 118 is configured to connect the system 100 with the one or more analytical devices, *e.g.*, via the communication network 104. In some embodiments, the one or more analytical devices include a laser ablation-inductively coupled plasma-mass spectrometer (LA-ICP-MS), a fluorescence image sensor, or a Raman spectrometer.

[0072] Although **FIG. 1** depicts a “system 100,” the figure is intended more as functional description of the various features which may be present in computer systems than as a structural schematic of the implementations described herein. In practice, and as recognized by those of ordinary skill in the art, items shown separately may be combined and some items may be separated. Moreover, although **FIG. 1** depicts certain data and modules in non-persistent memory 111, some or all of these data and modules may be in persistent memory 112.

[0073] In some embodiments, a method of the present disclosure comprises obtaining a biological sample (*e.g.*, a strand of hair including a hair shaft). The subject, in some embodiments, is a human. In some embodiments, the subject is a child aged equal to or below 12 years (*e.g.*, the child is aged equal to or below 5 years, 4 years, 3 years, 2 years, 1 year, 9 months, 6 months, 3 months, or 1 month). In some embodiments, the child is between the ages of about 12 and about 5 years old. In some embodiments, the subject is less than about 12, 11, 10, 9, 8, 7, 5, 4, 3, 2, or 1 year(s) old. In some embodiments, the subject is at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 year(s) old. In some embodiments, the subject is an adult. In some embodiments, the subject is an adult. **FIG. 2A** shows an example of a hair sample of a subject including a hair shaft. In some embodiments, the hair sample is cut from the subject (*e.g.*, with help of scissors). In some embodiments, the method of obtaining the hair sample is non-invasive. In some embodiments, the obtained hair sample has a minimum length of 1 cm (*e.g.*, the hair sample is 1 cm, 2 cm, 3 cm, 4 cm, or 5 cm long). In some embodiments, the obtained hair sample is at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 10, or at least 20 cm long. In some embodiments, the obtained hair sample is no more than 40, no more than 20, no more than 10, or no more than 5 cm long. In some embodiments, the obtained hair sample is from 1 to 5, from 4 to 10, from 8 to 20, or from 15 to 40 cm long. In some embodiments, the obtained hair sample falls within another range starting no lower than 1 cm and ending no higher than 40 cm long. In some embodiments, the hair sample includes any portion of a hair (*e.g.*, a tip or a portion between the tip and a follicle). In particular, there is no special requirement for the hair sample to include the

hair follicle. **FIG. 2B** shows an example of a tooth sample of a subject. **FIG. 2C** shows an example of a nail sample of a subject. In some embodiments, in instances of a nail or a hair, obtaining a biological sample refers to positioning the subject such that the nail or the hair is sampled. In some embodiments, the nail sample comprises a whole nail or a nail clipping.

[0074] In some embodiments, the obtained biological sample is pre-processed, such as being pre-treated by washing the biological sample with one or more solvents and/or surfactants and drying. In some embodiments where the biological sample is a hair, the hair sample is washed in a solution of TRITON X-100® and ultrapure metal free water (*e.g.*, MILLI-Q® water) and dried overnight in an oven (*e.g.*, at 60 degrees Celsius). In some embodiments, the pre-treatment further includes preparing the hair shaft for a measurement by placing the hair shaft on a glass slide (*e.g.*, a microscopic glass slide) with an adhesive film (*e.g.*, a double-sided tape). In some embodiments, the hair shaft is positioned such that the hair shaft is substantially straight. In some embodiments, the glass slide with the hair shaft is placed into or in the vicinity of a measurement system (*e.g.*, a laser ablation-inductively coupled plasma-mass spectrometer (LA-ICP-MS), a fluorescence image sensor, or a Raman spectrometer) for performing analysis. In some embodiments, where the biological sample is a tooth or a nail, a surface of the biological sample is cleaned (*e.g.*, by surfactant, water, or one or more solvents). In some embodiments, the sample is sectioned and then placed into or in the vicinity of a measurement system (*e.g.*, a laser ablation-inductively coupled plasma-mass spectrometer (LA-ICP-MS), a fluorescence image sensor, or a Raman spectrometer) for performing analysis.

[0075] **FIG. 3** shows a flow chart of a method 300 for evaluating a subject for a biological condition, such as a method for predicting a subject's diagnostic status with respect to a disease or disorder. In some embodiments, the method 300 comprises exposing a biological sample of the subject to a light source (as in operation 302). In some cases, the light source comprises a laser. In some embodiments, the analyzing determines temporal dynamics of underlying biological processes. In some embodiments, the analyzing comprises reducing a dimensionality of the plurality of Raman spectra (*e.g.*, by independent components analysis) prior to the processing. In some embodiments, the optical signal is generated by a light source (*e.g.*, a laser). In some embodiments, the biological sample comprises a tooth sample, a hair sample, or a nail sample. Next, in some embodiments, the method 300 comprises acquiring a plurality of Raman spectra from the exposed biological sample (as in operation 304). Next, in some embodiments, the method 300 comprises processing the plurality of Raman spectra to generate a spatial map of

the plurality of Raman spectra (as in operation 306). Next, in some embodiments, the method 300 comprises predicting a subject's diagnostic status with respect to a disease or disorder based at least in part on the spatial map of the plurality of Raman spectra (as in operation 308).

[0076] In some embodiments, the plurality of Raman spectra are acquired using a Raman spectroscopy microscope, including a 50X air coupled objective or a 63X water immersion coupled objection. In some embodiments, the laser comprises a wavelength of about 785 nm, or a wavelength of about 532 nm. In some embodiments, the acquiring is performed using an integration time of about 0.2 seconds to about 0.3 seconds. In some embodiments, the acquiring comprises moving the biological sample with a step size of about 2 microns to about 5 microns, subsequent to acquiring a Raman spectrum of the plurality of Raman spectra.

[0077] In some embodiments, the analyzing comprises generating a temporal Raman profile based at least in part on the Raman spectra acquired, and analyzing the temporal profile of variability in the Raman spectra. In some embodiments, at least a portion of the temporal Raman profile corresponds to a prenatal period of the subject.

[0078] In some embodiments, measurement data is collected from the biological sample sequentially at a plurality of positions along the biological sample. In some embodiments, the corresponding plurality of positions includes at least 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, 10000, or more than 10000 positions. In some embodiments, the respective positions are adjacent to each other. By this method, each area corresponding to a distinct position on the biological sample can be thereby associated with a dynamic (*e.g.*, time-varying) abundance measurement. In some embodiments, the respective positions are separated by a predefined distance. In some embodiments, the sampling is performed along the reference line of the biological sample starting from a respective position nearest to the tip of the biological sample such as hair sample (*e.g.*, at a position that corresponds to the youngest age of the subject). In general, the sampling can be performed starting from a respective position nearest to the tip or the root, as long as the direction of the sampling is known, and an appropriate trained model is used for the analyses.

[0079] In some embodiments, the sampling produces sets of data points. Each set of data points, in some embodiments, corresponds to a measurement (*e.g.*, an abundance or concentration) of a substance that is indicative of a dynamic biological response measured at a plurality of positions along the biological sample. In some embodiments, each position on the reference line of the biological sample corresponds to a specific time of growth of the biological sample. In some

embodiments, in an instance of the hair shaft, each position corresponds to approximately 20 min period of hair growth (*e.g.*, the period of hair growth calculated using a 5-micrometer laser step size and an average rate of hair growth 1 cm per month). By correlating the plurality of positions along the reference line of the biological sample to corresponding time periods of the growth, a first dataset including a plurality of traces is obtained. Each trace includes a time-dependent abundance of a measurement (*e.g.*, an abundance or concentration) of a substance that is indicative of a dynamic biological response measured from the biological sample. For example, the distance between positions may correspond to an estimated growth of the biological sample (*e.g.*, biological time). For example, abundance may be measured for a hair sample along a 1.2 cm distance, which corresponds to a biological time of approximately 35 days. In some embodiments, the biological time is estimated by using an average rate of hair growth (*e.g.*, 1 cm per month).

[0080] In some embodiments, data analysis of the Raman spectra comprises cosmic ray removal, background correction, spectral normalization, peak fitting, or any combination therein.

[0081] In some embodiments, data analysis is performed on the traces corresponding to a time-dependent abundance (*e.g.*, a time-dependent concentration) of a substance that is indicative of a dynamic biological response measured from the biological sample. This can comprise customized operations to clean the data (*e.g.*, smoothening the data over a time span, and/or removing data points that are higher or lower than a predetermined threshold). In some embodiments, the data analysis includes removing, from the traces, data points that have a mean absolute difference between adjacent data points that is at least one, two, or three times a standard deviation of the mean absolute difference between adjacent points.

[0082] In some embodiments, the data analysis further includes a dimension-reduction step, whereby the high-dimensional array of Raman spectra are decomposed into a lower dimensional array of derived time-varying components. Methods for dimensionality-reduction include independent component analysis (ICA), principal component analysis (PCA), non-negative matrix factorization (NNMF), and related unsupervised and supervised methods.

[0083] In some embodiments, the data analysis further includes performing recurrence quantification analysis (RQA) on the time-dependent traces, or on components derived from dimensionality-reduction techniques (ICA/PCA) applied to the time-dependent traces, to obtain a set of features that describe dynamical periodical characteristics of the traces. RQA measures variability in the time-dependent traces or components derived from the time-dependent traces.

RQA involves the estimation of features that describe periodic properties in a given waveform, which include the recurrence rates, determinism, mean diagonal length, maximum diagonal length, divergence, Shannon entropy in diagonal length, trend in recurrences, laminarity, trapping time, maximum vertical line length, Shannon entropy in vertical line lengths, mean recurrence time, Shannon entropy in recurrence times, number of the most probable recurrences, determination of a linear slope, determination of a plurality of non-linear parameters describing curvature of the one or more traces, determination of an abrupt change in intensity of the one or more traces, determination of one or more changes in a baseline intensity of the one or more traces, determination of a change of a frequency-domain representation of the one or more traces, determination of a change of the power-spectral domain representation of the one or more traces, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, or determination of a maximum Lyapunov exponent, and/or any combination thereof. Methods and features of RQA are described, for example, by Webber *et al.* in “Simpler Methods Do It Better: Success of Recurrence Quantification Analysis as a General Purpose Data Analysis Tool,” *Physics Letters A* 373, 3753-3756 (2009) and by Marwan *et al.* in “Recurrence Plots for the Analysis of Complex Systems,” *Physics Reports* 438, 237-239 (2007), the contents of each of which are herein incorporated by reference in their entirety. In some embodiments, the time-dependent traces are analyzed by using other analytical methods, such as Fourier Transformations, Wavelet Analysis, and Cosinor analysis. Such techniques can be applied to derive similar metrics, including spectral analysis of frequency components and their associated power. These metrics and associated derivative measures can be used in place of the features derived from RQA to analyze the time-dependent traces obtained from biological samples for purposes of predictive classification.

[0084] The RQA includes construction of recurrence plots that visualize and analyze dynamical temporal structures in respective obtained traces. Such recurrence plots can illustrate phasic processes in sequential measurements by plotting a given sequence against a time-lagged derivation of that sequence. From the one dimensional trace measured from the hair shaft, additional dimensions are computationally derived to embed the trace in a higher dimensional space referred to as a phase portrait, where t refers to the values of the original trace, and dimensions $(t+\tau)$ and $(t+2\tau)$ are derived from lagging the original time series by interval τ .

Subsequent analyses are then undertaken on the embedded phase portrait to construct recurrence plots and recurrence quantification analysis. A recurrence quantification plot can be derived from the phase portrait through the application of a threshold function to each point in the phase portrait; on the corresponding recurrence plot, consisting of a square binary matrix, typically represented as white or black space, a given point is assigned a value of 1 at each temporal interval wherein another point in the phase-portrait shares the spatial limits of the assigned threshold boundary. The RQA method is applied to the recurrence plot to examine the interval of delay between states in a given system, with a black point reflecting the temporal interval when a system revisits the same state. Periodic processes, where a system successively reiterates a given pattern of states, will manifest in a recurrence plot as diagonal black lines, whereas periods of stability will manifest as square structures, spurious repetitions as black dots, and unique events as white space.

[0085] In some embodiments, the recurrence plots are constructed for traces of a single substance or a combination of two substances (*e.g.*, in order to visualize an interactive periodic pattern of two substances; this can be referred to as cross-recurrence quantification analysis, or joint-recurrence quantification analysis). In some embodiments, the recurrence plots are constructed for a combination of three or more substances.

[0086] In some embodiments, the data analysis includes analyzing the recurrence plots to obtain a set of features associated with the recurrence plots. The features, which interchangeably can be termed “rhythmicity features,” or “dynamic features,” provide a quantitative measure describing the periodicity, predictability, and transitivity present in the plurality of traces. The features are selected from a set including recurrence rates, determinism, mean diagonal length, maximum diagonal length, divergence, Shannon entropy in diagonal length, trend in recurrences, laminarity, trapping time, maximum vertical line length, Shannon entropy in vertical line lengths, mean recurrence time, Shannon entropy in recurrence times, number of the most probable recurrences, determination of a linear slope, determination of a plurality of non-linear parameters describing curvature of the one or more traces, determination of an abrupt change in intensity of the one or more traces, determination of one or more changes in a baseline intensity of the one or more traces, determination of a change of a frequency-domain representation of the one or more traces, determination of a change of the power-spectral domain representation of the one or more traces, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more

joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, or determination of a maximum Lyapunov exponent, and/or any combination thereof.

[0087] In some embodiments, the data analysis further includes inputting the obtained set of features to a trained models. In some embodiments, the trained model includes a predictive computational algorithm to obtain a probability for the subject having a biological condition. In some embodiments, the predictive computational algorithm performs the following calculation:

$$p(\text{subject}) = \frac{1}{1 + e^{-(\alpha + \beta_1 x_1 + \dots + \beta_k x_k)}}$$

where $p(\text{subject})$ is the probability that the subject has the first biological condition, e is Euler's number, α is a calculated parameter associated with the probability that the subject has the biological condition when $\beta_1 x_1 + \dots + \beta_k x_k$ equals to zero, $x_1, \dots, x_k$ corresponds to a value derived for each feature in the set of features, the set of features including features from 1 through k , and $\beta_1, \dots, \beta_k$ corresponds to a weight parameter associated with each feature in the set of features including features from 1 through k .

[0088] In some embodiments, the weight parameters $\beta_1, \dots, \beta_k$ are defined based on model training. In some embodiments, the probability $p(\text{subject})$ is provided as a number ranging from 0 to 1, where 1 corresponds to a 100% probability that the subject has a biological condition.

[0089] In some embodiments, the data analysis includes applying a threshold to the obtained probability $p(\text{subject})$. If the obtained probability $p(\text{subject})$ is above the predetermined threshold, the subject is evaluated as having the biological condition. If the obtained probability is below the threshold, the subject is evaluated as not having the biological condition. In some embodiments, the threshold is between about 0.3 and 0.6 (e.g., the predetermined threshold is about 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, or 0.6). In some embodiments, the threshold is at least 0.1, at least 0.2, at least 0.3, at least 0.4, at least 0.5, at least 0.6, or at least 0.7. In some embodiments, the threshold is no more than 0.9, no more than 0.8, no more than 0.7, no more than 0.6, no more than 0.5, or no more than 0.4. In some embodiments, the threshold falls within another range starting no lower than 0.1 and ending no higher than 0.9. In some embodiments, the value assigned for a probabilistic threshold is predetermined, or estimated during the training of the model through the use of receiver-operating-characteristic (ROC) charts, with the optimal

threshold used corresponding to the value which yields the maximum area-under-the-curve (ROC-AUC). In some embodiments, the obtained probability is expressed in terms of associated odds (*e.g.*, odds ratio (OR), which can be derived from a probability such that $OR = p/(1-p)$). For example, the evaluation includes evaluating odds that the subject has the biological condition.

[0090] In some embodiments, the data analysis includes discriminating a first biological condition from an alternative condition, *e.g.*, a second, biological condition. In some embodiments, the alternative condition is associated with no known condition (*e.g.*, a neurotypical condition (NT)). In some embodiments, the first biological condition is associated with autism spectrum disorder (ASD) and the alternative condition is associated with an attention-deficit/hyperactivity disorder (ADHD). In some embodiments, the alternative condition is any other neurodevelopmental condition, or a comorbid diagnosis for two neurodevelopmental conditions. Therefore, in some embodiments, the data analysis is capable of discriminating between two neurodevelopmental conditions (*e.g.*, between autism spectrum disorder and ADHD, or between ASD and co-morbid (CM) cases diagnosed for both ASD and ADHD).

[0091] Health care providers, such as physicians and treating teams of a patient may have access to patient data (*e.g.*, dynamic biological response data or other health data), and/or predictions or assessments generated from such data. Based on the data analysis results, health care providers may determine clinical decisions or outcomes.

[0092] For example, a physician can instruct that patient undergo one or more clinical tests at the hospital or other clinical site, based at least in part on a predicted disease or disorder in the subject. In some embodiments, these instructions are provided when a certain pre-determined criterion is met (*e.g.*, a minimum threshold for a likelihood of the disease or disorder).

[0093] In some embodiments, such a minimum threshold includes, for example, at least about a 5% likelihood, at least about a 10% likelihood, at least about a 20% likelihood, at least about a 25% likelihood, at least about a 30% likelihood, at least about a 35% likelihood, at least about a 40% likelihood, at least about a 45% likelihood, at least about a 50% likelihood, at least about a 55% likelihood, at least about a 60% likelihood, at least about a 65% likelihood, at least about a 70% likelihood, at least about a 75% likelihood, at least about an 80% likelihood, at least about a 85% likelihood, at least about a 90% likelihood, at least about a 95% likelihood, at least about a 96% likelihood, at least about a 97% likelihood, at least about a 98% likelihood, or at least about a 99% likelihood. In some embodiments, the minimum threshold is no more than 99%, no more than 90%, no more than 80%, no more than 70%, no more than 60%, no more than

50%, or no more than 40%. In some embodiments, the minimum threshold is from 5% to 20%, from 10% to 50%, from 30% to 70%, or from 60% to 99%. In some embodiments, the minimum threshold falls within another range starting no lower than 5% and ending no higher than 99%.

[0094] As another example, a physician may prescribe a therapeutically effective dose of a treatment (*e.g.*, drug), a clinical procedure, or further clinical testing to be administered to the patient based at least in part on a predicted disease or disorder in the subject. For example, the physician may prescribe an anti-inflammatory therapeutic in response to an indication of inflammation in the patient.

[0095] **Models**

[0096] In some embodiments, the methods and systems of the present disclosure utilize or access external capabilities of artificial intelligence techniques to develop signatures for various diseases or disorders. In some embodiments, these signatures are used to accurately predict diseases or disorders (*e.g.*, months or years earlier than with standard of clinical care). Using such a predictive capability, health care providers (*e.g.*, physicians) are able to make informed, accurate risk-based decisions, thereby improving quality of care and monitoring provided to patients.

[0097] In some embodiments, the methods and systems of the present disclosure analyze acquired dynamic biological response data from a subject (patient) to generate a likelihood of the subject having a disease or disorder. For example, in some embodiments, the system applies a trained (*e.g.*, prediction) algorithm to the acquired dynamic biological response data to generate the likelihood of the subject having a disease or disorder. In some embodiments, the trained algorithm comprises an artificial intelligence-based model, such as a machine learning based classifier, configured to process the acquired dynamic biological response data to generate the likelihood of the subject having the disease or disorder. In some embodiments, the model is trained using clinical datasets from one or more cohorts of patients, *e.g.*, using clinical health data and/or dynamic biological response data of the patients as inputs and known clinical health outcomes (*e.g.*, disease or disorder) of the patients as outputs to the model.

[0098] In some embodiments, the model comprises one or more machine learning algorithms. Examples of machine learning algorithms include, but are not limited to, a support vector machine (SVM), a naïve Bayes classification, a random forest, a neural network (such as a deep neural network (DNN), a recurrent neural network (RNN), a deep RNN, a long short-term memory (LSTM) recurrent neural network (RNN), a gated recurrent unit (GRU), or other

supervised learning algorithm or unsupervised machine learning, statistical, or deep learning algorithm for classification and regression. In some embodiments, the model likewise involves the estimation of ensemble models, comprised of multiple predictive models, and utilize techniques such as gradient boosting, for example in the construction of gradient-boosting decision trees. In some embodiments, the model is trained using one or more training datasets corresponding to patient data.

[0099] In some embodiments, training datasets are generated from, for example, one or more cohorts of patients having common clinical characteristics (features) and clinical outcomes (labels). In some embodiments, training datasets comprise a set of features and labels corresponding to the features. In some embodiments, features correspond to algorithm inputs comprising dynamic biological response data, patient demographic information derived from electronic medical records (EMR), and medical observations. In some embodiments, features comprise clinical characteristics such as, for example, certain ranges or categories of dynamic biological response data. In some embodiments, features comprise patient information such as patient age, patient medical history, other medical conditions, current or past medications, and time since the last observation. For example, a set of features collected from a given patient at a given time point may collectively serve as a signature, which may be indicative of a health state or status of the patient at the given time point.

[00100] For example, in some embodiments, ranges of dynamic biological response data and other health measurements are expressed as a plurality of disjoint continuous ranges of continuous measurement values, and categories of dynamic biological response data and other health measurements may be expressed as a plurality of disjoint sets of measurement values (*e.g.*, {"high", "low"}, {"high", "normal"}, {"low", "normal"}, {"high", "borderline high", "normal", "low"}, *etc.*). In some embodiments, clinical characteristics include clinical labels indicating the patient's health history, such as a diagnosis of a disease or disorder, a previous administration of a clinical treatment (*e.g.*, a drug, a surgical treatment, chemotherapy, radiotherapy, immunotherapy, *etc.*), behavioral factors, or other health status (*e.g.*, hypertension or high blood pressure, hyperglycemia or high blood glucose, hypercholesterolemia or high blood cholesterol, history of allergic reaction or other adverse reaction, *etc.*).

[00101] In some embodiments, labels comprise clinical outcomes such as, for example, a presence, absence, diagnosis, or prognosis of a disease or disorder in the subject (*e.g.*, patient). In some embodiments, clinical outcomes include a temporal characteristic associated with the

presence, absence, diagnosis, or prognosis of the disease or disorder in the patient. For example, temporal characteristics may be indicative of the patient having had an occurrence of the disease or disorder within a certain period of time after a previous clinical outcome (*e.g.*, being discharged from the hospital, being administered a treatment such as medication, undergoing a clinical procedure such as surgical operation, *etc.*). In some embodiments, such a period of time includes, for example, about 1 hour, about 2 hours, about 3 hours, about 4 hours, about 6 hours, about 8 hours, about 10 hours, about 12 hours, about 14 hours, about 16 hours, about 18 hours, about 20 hours, about 22 hours, about 24 hours, about 2 days, about 3 days, about 4 days, about 5 days, about 6 days, about 7 days, about 10 days, about 2 weeks, about 3 weeks, about 4 weeks, about 1 month, about 2 months, about 3 months, about 4 months, about 6 months, about 8 months, about 10 months, about 1 year, or more than about 1 year. In some embodiments, the period of time is no more than 5 years, no more than 1 year, no more than 6 months, no more than 3 months, no more than 1 month, no more than 2 weeks, no more than 1 week, no more than 1 day, or no more than 12 hours. In some embodiments, the period of time is from 1 hour to 12 hours, from 12 hours to 24 hours, from 1 day to 7 days, from 1 week to 4 weeks, from 1 month to 12 months, or from 1 year to 5 years. In some embodiments, the period of time falls within another range starting no lower than 1 hour and ending no higher than 5 years.

[00102] In some embodiments, input features are structured by aggregating the data into bins or alternatively using a one-hot encoding. In some embodiments, inputs also include feature values or vectors derived from the previously mentioned inputs, such as cross-correlations calculated between separate dynamic biological response data or other measurements over a fixed period of time, and the discrete derivative or the finite difference between successive measurements. In some embodiments, such a period of time includes, for example, about 1 hour, about 2 hours, about 3 hours, about 4 hours, about 6 hours, about 8 hours, about 10 hours, about 12 hours, about 14 hours, about 16 hours, about 18 hours, about 20 hours, about 22 hours, about 24 hours, about 2 days, about 3 days, about 4 days, about 5 days, about 6 days, about 7 days, about 10 days, about 2 weeks, about 3 weeks, about 4 weeks, about 1 month, about 2 months, about 3 months, about 4 months, about 6 months, about 8 months, about 10 months, about 1 year, or more than about 1 year. In some embodiments, the period of time is no more than 5 years, no more than 1 year, no more than 6 months, no more than 3 months, no more than 1 month, no more than 2 weeks, no more than 1 week, no more than 1 day, or no more than 12 hours. In some embodiments, the period of time is from 1 hour to 12 hours, from 12 hours to 24 hours, from 1

day to 7 days, from 1 week to 4 weeks, from 1 month to 12 months, or from 1 year to 5 years. In some embodiments, the period of time falls within another range starting no lower than 1 hour and ending no higher than 5 years.

[00103] In some embodiments, training records are constructed from sequences of observations. In some embodiments, such sequences comprise a fixed length for ease of data processing. For example, sequences may be zero-padded or selected as independent subsets of a single patient's records.

[00104] In some embodiments, the model processes the input features to generate output values comprising one or more classifications, one or more predictions, or a combination thereof. For example, such classifications or predictions may include a binary classification of a healthy/normal health state (*e.g.*, absence of a disease or disorder) or an adverse health state (*e.g.*, presence of a disease or disorder), a classification between a group of categorical labels (*e.g.*, 'no disease or disorder', 'apparent disease or disorder', and 'likely disease or disorder'), a likelihood (*e.g.*, relative likelihood or probability) of developing a particular disease or disorder, a score indicative of a presence of disease or disorder, a score indicative of a level of systemic inflammation experienced by the patient, a 'risk factor' for the likelihood of mortality of the patient, a prediction of the time at which the patient is expected to have developed the disease or disorder, and a confidence interval for any numeric predictions. In some embodiments, various machine learning techniques are cascaded such that the output of a machine learning technique is also used as input features to subsequent layers or subsections of the model.

[00105] In order to train the model (*e.g.*, by determining weights and correlations of the model) to generate real-time classifications or predictions, the model can be trained using datasets. In some embodiments, such datasets are sufficiently large to generate statistically significant classifications or predictions. For example, datasets may comprise: databases of de-identified data including dynamic biological response data and other measurements, and dynamic biological response data and other measurements from a hospital or other clinical setting.

[00106] In some embodiments, datasets are split into subsets (*e.g.*, discrete or overlapping), such as a training dataset, a development dataset, and a test dataset. For example, a dataset may be split into a training dataset comprising 80% of the dataset, a development dataset comprising 10% of the dataset, and a test dataset comprising 10% of the dataset. In some embodiments, the training dataset comprises about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, or about 90% of the dataset. In some embodiments, the

development dataset comprises about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, or about 90% of the dataset. In some embodiments, the test dataset comprises about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, or about 90% of the dataset. In some embodiments, training sets (*e.g.*, training datasets) are selected by random sampling of a set of data corresponding to one or more patient cohorts to ensure independence of sampling. Alternatively, in some embodiments, training sets (*e.g.*, training datasets) are selected by proportionate sampling of a set of data corresponding to one or more patient cohorts to ensure independence of sampling.

[00107] In some embodiments, to improve the accuracy of model predictions and reduce overfitting of the model, the datasets are augmented to increase the number of samples within the training set. For example, data augmentation may comprise rearranging the order of observations in a training record. To accommodate datasets having missing observations, methods to impute missing data may be used, such as forward-filling, back-filling, linear interpolation, and multi-task Gaussian processes. Datasets may be filtered to remove confounding factors. For example, within a database, a subset of patients may be excluded.

[00108] The model may comprise one or more neural networks, such as a neural network, a convolutional neural network (CNN), a deep neural network (DNN), a recurrent neural network (RNN), or a deep RNN. The recurrent neural network may comprise units which can be long short-term memory (LSTM) units or gated recurrent units (GRU). For example, the model may comprise an algorithm architecture comprising a neural network with a set of input features such as vital sign and other measurements, patient medical history, and/or patient demographics. Neural network techniques, such as dropout or regularization, may be used during training the model to prevent overfitting. The neural network may comprise a plurality of sub-networks, each of which is configured to generate a classification or prediction of a different type of output information (*e.g.*, which may be combined to form an overall output of the neural network). The machine learning model may alternatively utilize statistical or related algorithms including random forest, classification and regression trees, support vector machines, discriminant analyses, regression techniques, as well as ensemble and gradient-boosted variations thereof.

[00109] When the model generates a classification or a prediction of a disease or disorder, a notification (*e.g.*, alert or alarm) may be generated and transmitted to a health care provider, such as a physician, nurse, or other member of the patient's treating team within a hospital. Notifications may be transmitted via an automated phone call, a short message service (SMS) or

multimedia message service (MMS) message, an e-mail, or an alert within a dashboard. The notification may comprise output information such as a prediction of a disease or disorder, a likelihood of the predicted disease or disorder, a time until an expected onset of the disease or disorder, a confidence interval of the likelihood or time, or a recommended course of treatment for the disease or disorder.

[00110] To validate the performance of the model, different performance metrics may be generated. For example, an area under the receiver-operating curve (AUROC) may be used to determine the diagnostic capability of the model. For example, the model may use classification thresholds which are adjustable, such that specificity and sensitivity are tunable, and the receiver-operating curve (ROC) can be used to identify the different operating points corresponding to different values of specificity and sensitivity.

[00111] In some cases, such as when datasets are not sufficiently large, cross-validation may be performed to assess the robustness of a model across different training and testing datasets.

[00112] To calculate performance metrics such as sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), AUPRC, AUROC, or similar, the following definitions may be used. A “false positive” may refer to an outcome in which a positive outcome or result has been incorrectly or prematurely generated (*e.g.*, before the actual onset of, or without any onset of, the disease or disorder). A “true positive” may refer to an outcome in which positive outcome or result has been correctly generated, when the patient has the disease or disorder (*e.g.*, the patient shows symptoms of the disease or disorder, or the patient’s record indicates the disease or disorder). A “false negative” may refer to an outcome in which a negative outcome or result has been generated, but the patient has the disease or disorder (*e.g.*, the patient shows symptoms of the disease or disorder, or the patient’s record indicates the disease or disorder). A “true negative” may refer to an outcome in which a negative outcome or result has been generated (*e.g.*, before the actual onset of, or without any onset of, the disease or disorder).

[00113] The model may be trained until certain pre-determined conditions for accuracy or performance are satisfied, such as having minimum desired values corresponding to diagnostic accuracy measures. For example, the diagnostic accuracy measure may correspond to prediction of a likelihood of occurrence of a disease or disorder in the subject. As another example, the diagnostic accuracy measure may correspond to prediction of a likelihood of deterioration or recurrence of a disease or disorder for which the subject has previously been treated. Examples of

diagnostic accuracy measures may include sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, area under the precision-recall curve (AUPRC), and area under the curve (AUC) of a Receiver Operating Characteristic (ROC) curve (AUROC) corresponding to the diagnostic accuracy of detecting or predicting a disease or disorder.

[00114] For example, such a pre-determined condition is, in some embodiments, that the sensitivity of predicting the disease or disorder comprises a value of, for example, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%. In some embodiments, the pre-determined condition is that the sensitivity comprises a value of no more than 100%, no more than 99%, no more than 90%, no more than 80%, no more than 70%, or no more than 60%. In some embodiments, the pre-determined condition is that the sensitivity comprises a value from 50% to 70%, from 60% to 80%, from 70% to 90%, or from 90% to 100%. In some embodiments, the pre-determined condition is that the sensitivity comprises a value that falls within another range starting no lower than 50% and ending no higher than 100%.

[00115] As another example, such a pre-determined condition is, in some embodiments, that the specificity of predicting the disease or disorder comprises a value of, for example, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%. In some embodiments, the pre-determined condition is that the specificity comprises a value of no more than 100%, no more than 99%, no more than 90%, no more than 80%, no more than 70%, or no more than 60%. In some embodiments, the pre-determined condition is that the specificity comprises a value from 50% to 70%, from 60% to 80%, from 70% to 90%, or from 90% to 100%. In some embodiments, the pre-determined condition is that the specificity comprises a value that falls within another range starting no lower than 50% and ending no higher than 100%.

[00116] As another example, such a pre-determined condition is, in some embodiments, that the positive predictive value (PPV) of predicting the disease or disorder comprises a value of, for example, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%. In

some embodiments, the pre-determined condition is that the PPV comprises a value of no more than 100%, no more than 99%, no more than 90%, no more than 80%, no more than 70%, or no more than 60%. In some embodiments, the pre-determined condition is that the PPV comprises a value from 50% to 70%, from 60% to 80%, from 70% to 90%, or from 90% to 100%. In some embodiments, the pre-determined condition is that the PPV comprises a value that falls within another range starting no lower than 50% and ending no higher than 100%.

[00117] As another example, such a pre-determined condition is, in some embodiments, that the negative predictive value (NPV) of predicting the disease or disorder comprises a value of, for example, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%. In some embodiments, the pre-determined condition is that the NPV comprises a value of no more than 100%, no more than 99%, no more than 90%, no more than 80%, no more than 70%, or no more than 60%. In some embodiments, the pre-determined condition is that the NPV comprises a value from 50% to 70%, from 60% to 80%, from 70% to 90%, or from 90% to 100%. In some embodiments, the pre-determined condition is that the NPV comprises a value that falls within another range starting no lower than 50% and ending no higher than 100%.

[00118] As another example, such a pre-determined condition is, in some embodiments, that the area under the curve (AUC) of a Receiver Operating Characteristic (ROC) curve (AUROC) of predicting the disease or disorder comprises a value of at least about 0.50, at least about 0.55, at least about 0.60, at least about 0.65, at least about 0.70, at least about 0.75, at least about 0.80, at least about 0.85, at least about 0.90, at least about 0.95, at least about 0.96, at least about 0.97, at least about 0.98, or at least about 0.99. In some embodiments, the pre-determined condition is that the AUROC comprises a value of no more than 1, no more than 0.99, no more than 0.90, no more than 0.80, no more than 0.70, or no more than 0.60. In some embodiments, the pre-determined condition is that the AUROC comprises a value from 0.50 to 0.70, from 0.60 to 0.80, from 0.70 to 0.90, or from 0.90 to 1. In some embodiments, the pre-determined condition is that the AUROC comprises a value that falls within another range starting no lower than 0.50 and ending no higher than 1.

[00119] As another example, such a pre-determined condition is, in some embodiments, that the area under the precision-recall curve (AUPRC) of predicting the disease or disorder comprises a value of at least about 0.10, at least about 0.15, at least about 0.20, at least about

0.25, at least about 0.30, at least about 0.35, at least about 0.40, at least about 0.45, at least about 0.50, at least about 0.55, at least about 0.60, at least about 0.65, at least about 0.70, at least about 0.75, at least about 0.80, at least about 0.85, at least about 0.90, at least about 0.95, at least about 0.96, at least about 0.97, at least about 0.98, or at least about 0.99. In some embodiments, the pre-determined condition is that the AUPRC comprises a value of no more than 1, no more than 0.99, no more than 0.90, no more than 0.80, no more than 0.70, no more than 0.60, or no more than 0.50. In some embodiments, the pre-determined condition is that the AUPRC comprises a value of from 0.10 to 0.40, from 0.30 to 0.70, from 0.60 to 0.90, or from 0.80 to 1. In some embodiments, the pre-determined condition is that the AUPRC comprises a value that falls within another range starting no lower than 0.10 and ending no higher than 1.

[00120] In some embodiments, the trained model is trained or configured to predict the disease or disorder with a sensitivity of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%. In some embodiments, the model is trained or configured to predict the disease or disorder with a sensitivity of no more than 100%, no more than 99%, no more than 90%, no more than 80%, no more than 70%, or no more than 60%. In some embodiments, the model is trained or configured to predict the disease or disorder with a sensitivity of from 50% to 70%, from 60% to 80%, from 70% to 90%, or from 90% to 100%. In some embodiments, the model is trained or configured to predict the disease or disorder with a sensitivity that falls within another range starting no lower than 50% and ending no higher than 100%.

[00121] In some embodiments, the trained model is trained or configured to predict the disease or disorder with a specificity of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%. In some embodiments, the model is trained or configured to predict the disease or disorder with a specificity of no more than 100%, no more than 99%, no more than 90%, no more than 80%, no more than 70%, or no more than 60%. In some embodiments, the model is trained or configured to predict the disease or disorder with a specificity of from 50% to 70%, from 60% to 80%, from 70% to 90%, or from 90% to 100%. In some embodiments, the model is trained or configured to predict the disease or disorder with a specificity that falls within another range starting no lower than 50% and ending no higher than 100%.

[00122] In some embodiments, the trained model is trained or configured to predict the disease or disorder with a positive predictive value (PPV) of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%. In some embodiments, the model is trained or configured to predict the disease or disorder with a PPV of no more than 100%, no more than 99%, no more than 90%, no more than 80%, no more than 70%, or no more than 60%. In some embodiments, the model is trained or configured to predict the disease or disorder with a PPV of from 50% to 70%, from 60% to 80%, from 70% to 90%, or from 90% to 100%. In some embodiments, the model is trained or configured to predict the disease or disorder with a PPV that falls within another range starting no lower than 50% and ending no higher than 100%.

[00123] In some embodiments, the trained model is trained or configured to predict the disease or disorder with a negative predictive value (NPV) of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%. In some embodiments, the model is trained or configured to predict the disease or disorder with a NPV of no more than 100%, no more than 99%, no more than 90%, no more than 80%, no more than 70%, or no more than 60%. In some embodiments, the model is trained or configured to predict the disease or disorder with a NPV of from 50% to 70%, from 60% to 80%, from 70% to 90%, or from 90% to 100%. In some embodiments, the model is trained or configured to predict the disease or disorder with a NPV that falls within another range starting no lower than 50% and ending no higher than 100%.

[00124] In some embodiments, the trained model is trained or configured to predict the disease or disorder with an area under the curve (AUC) of a Receiver Operating Characteristic (ROC) curve (AUROC) of at least about 0.50, at least about 0.55, at least about 0.60, at least about 0.65, at least about 0.70, at least about 0.75, at least about 0.80, at least about 0.85, at least about 0.90, at least about 0.95, at least about 0.96, at least about 0.97, at least about 0.98, or at least about 0.99. In some embodiments, the trained model is trained or configured to predict the disease or disorder with an AUROC of no more than 1, no more than 0.99, no more than 0.90, no more than 0.80, no more than 0.70, or no more than 0.60. In some embodiments, the trained model is trained or configured to predict the disease or disorder with an AUROC of from 0.50 to 0.70, from 0.60 to 0.80, from 0.70 to 0.90, or from 0.90 to 1. In some embodiments, the trained model

is trained or configured to predict the disease or disorder with an AUROC that falls within another range starting no lower than 0.50 and ending no higher than 1.

[00125] In some embodiments, the trained model is trained or configured to predict the disease or disorder with an area under the precision-recall curve (AUPRC) of at least about 0.10, at least about 0.15, at least about 0.20, at least about 0.25, at least about 0.30, at least about 0.35, at least about 0.40, at least about 0.45, at least about 0.50, at least about 0.55, at least about 0.60, at least about 0.65, at least about 0.70, at least about 0.75, at least about 0.80, at least about 0.85, at least about 0.90, at least about 0.95, at least about 0.96, at least about 0.97, at least about 0.98, or at least about 0.99. In some embodiments, the model is trained or configured to predict the disease or disorder with an AUPRC of no more than 1, no more than 0.99, no more than 0.90, no more than 0.80, no more than 0.70, no more than 0.60, or no more than 0.50. In some embodiments, the model is trained or configured to predict the disease or disorder with an AUPRC of from 0.10 to 0.40, from 0.30 to 0.70, from 0.60 to 0.90, or from 0.80 to 1. In some embodiments, the model is trained or configured to predict the disease or disorder with an AUPRC that falls within another range starting no lower than 0.10 and ending no higher than 1.

[00126] In some embodiments, the training data sets are collected from training subjects (*e.g.*, humans). Each training has a diagnostic status indicating that they have either been diagnosed with the biological condition, or have not been diagnosed with the biological condition. In some embodiments, the training subjects are children aged equal to, or below, 12 years (*e.g.*, equal to or below 5 years, 4 years, 3 years, 2 years, 1 year, 9 months, 6 months, 3 months or 1 month). In some embodiments, the child is between the ages of about 12 and about 5 years old. In some embodiments, the subject is less than about 12, 11, 10, 9, 8, 7, 5, 4, 3, 2, or 1 year(s) old. In some embodiments, the subject is at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 year(s) old. In some embodiments, the following training procedure is performed for each training subject in a plurality of training subjects. In some embodiments, the training subjects is no more than 18, no more than 15, no more than 12, no more than 11, no more than 10, no more than 9, no more than 8, no more than 7, no more than 5, no more than 4, no more than 3, no more than 2, or no more than 1 year(s) old.

[00127] In some embodiments, a plurality of positions of a reference line on a biological sample of the training subject is sampled, thereby obtaining a plurality of dynamic biological response samples. Each dynamic biological response sample in the corresponding plurality of dynamic biological response samples for a different position in the corresponding plurality of positions, and

each position in the corresponding plurality of positions representing a different period of growth of the corresponding biological sample. Next, each respective dynamic biological response sample is analyzed (*e.g.*, using a laser ablation-inductively coupled plasma-mass spectrometer (LA-ICP-MS), a fluorescence image sensor, or a Raman spectrometer) to obtain a plurality of traces. Each trace in the corresponding plurality of traces corresponds to an abundance measurement of a corresponding substance, which are over time collectively determined from the corresponding plurality of dynamic biological response samples.

[00128] Next, a respective second dataset is obtained from the corresponding plurality of traces that includes a corresponding set of features, each respective feature in the corresponding set of features being determined by a variation of abundance of one or more substances in the corresponding plurality of traces as assessed by the application of recurrence quantification analysis or related methods to either the Raman waveform or dimensions derived from the Raman waveform through ICA/PCA or related dimensionality-reduction techniques.

[00129] Next, an untrained or partially untrained model is generated, with (i) the corresponding set of features of each respective second dataset of each training subject in the plurality of training subjects and (ii) the corresponding diagnostic status of each training subject in the plurality of training subjects, selected from among the first diagnostic status and the second diagnostic status, thereby obtaining a trained model. The trained model provides an indication as to whether a test subject has the first biological condition based on values for features in a set of features acquired from a biological sample of the test subject. In some embodiments, the trained model is a neural network algorithm, a support vector machine algorithm, a decision tree algorithm, an unsupervised clustering algorithm, a supervised clustering algorithm, a regression algorithm, or any combination or variant thereof, particularly including gradient-boosting implementations of the described algorithms, *e.g.*, gradient-boosted decision trees. In some embodiments, the trained machine learning model utilizes a gradient-boosted ensemble algorithm. In some embodiments, the trained model is a multinomial or a binomial classifier. In some embodiments, the trained model can be used to make a binary prediction as to whether a sample was derived from a subject with the first biological condition or not; or may be multinomial, distinguishing subjects with no diagnosis from those with the first biological condition or a second biological condition, where the second biological condition is distinct from the first biological condition.

[00130] In some embodiments, the model is a neural network or a convolutional neural network. See, Vincent *et al.*, 2010, “Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion,” J Mach Learn Res 11, pp. 3371-3408; Larochelle *et al.*, 2009, “Exploring strategies for training deep neural networks,” J Mach Learn Res 10, pp. 1-40; and Hassoun, 1995, Fundamentals of Artificial Neural Networks, Massachusetts Institute of Technology, each of which is hereby incorporated by reference.

[00131] Independent component analysis (ICA), such as that described herein in the unsupervised dimensionality-reduction of Raman waveforms, is described in Lee, T.-W. (1998): Independent component analysis: Theory and applications, Boston, Mass: Kluwer Academic Publishers, ISBN 0-7923-8261-7, and Hyvärinen, A.; Karhunen, J.; Oja, E. (2001): *Independent Component Analysis*, New York: Wiley, ISBN 978-0-471-40540-5, which is hereby incorporated by reference in its entirety.

[00132] Principal component analysis (PCA), such as that described herein in the unsupervised dimensionality-reduction of Raman waveforms, is described in Jolliffe, I. T. (2002). Principal Component Analysis. Springer Series in Statistics. New York: Springer-Verlag. doi:10.1007/b98835. ISBN 978-0-387-95442-4, which is hereby incorporated by reference in its entirety.

[00133] SVMs are described in Cristianini and Shawe-Taylor, 2000, “An Introduction to Support Vector Machines,” Cambridge University Press, Cambridge; Boser *et al.*, 1992, “A training algorithm for optimal margin classifiers,” in Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, ACM Press, Pittsburgh, Pa., pp. 142-152; Vapnik, 1998, *Statistical Learning Theory*, Wiley, New York; Mount, 2001, *Bioinformatics: sequence and genome analysis*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.; Duda, *Pattern Classification*, Second Edition, 2001, John Wiley & Sons, Inc., pp. 259, 262-265; and Hastie, 2001, *The Elements of Statistical Learning*, Springer, New York; and Furey *et al.*, 2000, *Bioinformatics* 16, 906-914, each of which is hereby incorporated by reference in its entirety. When used for classification, SVMs separate a given set of binary labeled data with a hyper-plane that is maximally distant from the labeled data. For cases in which no linear separation is possible, SVMs can work in combination with the technique of ‘kernels’, which automatically realizes a non-linear mapping to a feature space. The hyper-plane found by the SVM in feature space corresponds to a non-linear decision boundary in the input space.

[00134] Decision trees are described generally by Duda, 2001, *Pattern Classification*, John Wiley & Sons, Inc., New York, pp. 395-396, which is hereby incorporated by reference. Tree-based methods partition the feature space into a set of rectangles, and then fit a model (like a constant) in each one. In some embodiments, the decision tree is random forest regression. One specific algorithm that can be used is a classification and regression tree (CART). Other specific decision tree algorithms include, but are not limited to, ID3, C4.5, MART, and Random Forests. CART, ID3, and C4.5 are described in Duda, 2001, *Pattern Classification*, John Wiley & Sons, Inc., New York, pp. 396-408 and pp. 411-412, which is hereby incorporated by reference. CART, MART, and C4.5 are described in Hastie *et al.*, 2001, *The Elements of Statistical Learning*, Springer-Verlag, New York, Chapter 9, which is hereby incorporated by reference in its entirety. Random Forests are described in Breiman, 1999, “Random Forests—Random Features,” Technical Report 567, Statistics Department, U.C. Berkeley, September 1999, which is hereby incorporated by reference in its entirety.

[00135] Clustering (*e.g.*, unsupervised clustering model algorithms and supervised clustering model algorithms) is described at pages 211-256 of Duda and Hart, *Pattern Classification and Scene Analysis*, 1973, John Wiley & Sons, Inc., New York, (hereinafter “Duda 1973”) which is hereby incorporated by reference in its entirety. As described in Section 6.7 of Duda 1973, the clustering problem is described as one of finding natural groupings in a dataset. To identify natural groupings, two issues are addressed. First, a way to measure similarity (or dissimilarity) between two samples is determined. This metric (similarity measure) is used to ensure that the samples in one cluster are more like one another than they are to samples in other clusters. Second, a mechanism for partitioning the data into clusters using the similarity measure is determined. Similarity measures are discussed in Section 6.7 of Duda 1973, where it is stated that one way to begin a clustering investigation is to define a distance function and to compute the matrix of distances between all pairs of samples in the training set. If distance is a good measure of similarity, then the distance between reference entities in the same cluster will be significantly less than the distance between the reference entities in different clusters. However, as stated on page 215 of Duda 1973, clustering does not require the use of a distance metric. For example, a nonmetric similarity function $s(x, x')$ can be used to compare two vectors x and x' . Conventionally, $s(x, x')$ is a symmetric function whose value is large when x and x' are somehow “similar.” An example of a nonmetric similarity function $s(x, x')$ is provided on page 218 of Duda 1973. Once a method for measuring “similarity” or “dissimilarity” between points in a dataset has been selected,

clustering requires a criterion function that measures the clustering quality of any partition of the data. Partitions of the data set that extremize the criterion function are used to cluster the data. *See* page 217 of Duda 1973. Criterion functions are discussed in Section 6.8 of Duda 1973. More recently, Duda *et al.*, Pattern Classification, 2nd edition, John Wiley & Sons, Inc. New York, has been published. Pages 537-563 describe clustering in detail. More information on clustering techniques can be found in Kaufman and Rousseeuw, 1990, *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York, N.Y.; Everitt, 1993, *Cluster analysis* (3d ed.), Wiley, New York, N.Y.; and Backer, 1995, *Computer-Assisted Reasoning in Cluster Analysis*, Prentice Hall, Upper Saddle River, New Jersey, each of which is hereby incorporated by reference. Particular exemplary clustering techniques that can be used in the present disclosure include, but are not limited to, hierarchical clustering (agglomerative clustering using nearest-neighbor algorithm, farthest-neighbor algorithm, the average linkage algorithm, the centroid algorithm, or the sum-of-squares algorithm), k-means clustering, fuzzy k-means clustering algorithm, and Jarvis-Patrick clustering. In some embodiments, the clustering comprises unsupervised clustering, where no preconceived notion of what clusters should form when the training set is clustered, are imposed.

[00136] Regression models, such as that of the multi-category logit models, are described in Agresti, *An Introduction to Categorical Data Analysis*, 1996, John Wiley & Sons, Inc., New York, Chapter 8, which is hereby incorporated by reference in its entirety. In some embodiments, the model makes use of a regression model disclosed in Hastie *et al.*, 2001, *The Elements of Statistical Learning*, Springer-Verlag, New York, which is hereby incorporated by reference in its entirety. In some embodiments, gradient-boosting models are used toward, for example, the classification algorithms described herein; these gradient-boosting models are described in Boehmke, Bradley; Greenwell, Brandon (2019). "Gradient Boosting". *Hands-On Machine Learning with R*. Chapman & Hall. pp. 221–245. ISBN 978-1-138-49568-5., which is hereby incorporated by reference in its entirety. In some embodiments, ensemble modeling techniques are used, for example, toward the classification algorithms described herein; these ensemble modeling techniques are described in the implementation of classification models herein, are described in Zhou Zhihua (2012). *Ensemble Methods: Foundations and Algorithms*. Chapman and Hall/CRC. ISBN 978-1-439-83003-1, which is hereby incorporated by reference in its entirety.

[00137] In some embodiments, the machine learning analysis is performed by a device executing one or more programs (*e.g.*, one or more programs stored in the Non-Persistent Memory 111 or in the Persistent Memory 112 in Figure 1) including instructions to perform the data analysis. In some embodiments, the data analysis is performed by a system comprising at least one processor (*e.g.*, the processing core 102) and memory (*e.g.*, one or more programs stored in the Non-Persistent Memory 111 or in the Persistent Memory 112) comprising instructions to perform the data analysis.

[00138] **Computer systems**

[00139] The present disclosure provides computer systems that are programmed to implement methods of the disclosure. **FIG. 4** shows a computer system 401 that is programmed or otherwise configured to, for example, obtain a Raman signature of tooth samples, analyze the Raman spectra spatially across tooth samples, generate a temporal Raman profile, process data using trained models, and predict a subject's diagnostic status with respect to a disease or disorder. The computer system 401 can regulate various aspects of sensor data analysis of the present disclosure, such as, for example, staining a tooth sample, obtaining a fluorescence image of stained tooth samples, analyzing a fluorescence intensity spatially across stained tooth samples, generating a temporal Raman profile, measuring the dynamics of the temporal profile, process data using trained models, and predicting a subject's diagnostic status with respect to a disease or disorder. The computer system 401 can be an electronic device of a user or a computer system that is remotely located with respect to the electronic device. The electronic device can be a mobile electronic device.

[00140] The computer system 401 includes a central processing unit (CPU, also "processor" and "computer processor" herein) 405, which can be a single core or multi core processor, or a plurality of processors for parallel processing. The computer system 401 also includes memory or memory location 410 (*e.g.*, random-access memory, read-only memory, flash memory), electronic storage unit 415 (*e.g.*, hard disk), communication interface 420 (*e.g.*, network adapter) for communicating with one or more other systems, and peripheral devices 425, such as cache, other memory, data storage and/or electronic display adapters. The memory 410, storage unit 415, interface 420 and peripheral devices 425 are in communication with the CPU 405 through a communication bus (solid lines), such as a motherboard. The storage unit 415 can be a data storage unit (or data repository) for storing data. The computer system 401 can be operatively coupled to a computer network ("network") 430 with the aid of the communication interface 420.

The network 430 can be the Internet, an internet and/or extranet, or an intranet and/or extranet that is in communication with the Internet. The network 430 in some cases is a telecommunication and/or data network. The network 430 can include one or more computer servers, which can enable distributed computing, such as cloud computing. The network 430, in some cases with the aid of the computer system 401, can implement a peer-to-peer network, which may enable devices coupled to the computer system 401 to behave as a client or a server.

[00141] The CPU 405 can execute a sequence of machine-readable instructions, which can be embodied in a program or software. The instructions may be stored in a memory location, such as the memory 410. The instructions can be directed to the CPU 405, which can subsequently program or otherwise configure the CPU 405 to implement methods of the present disclosure. Examples of operations performed by the CPU 405 can include fetch, decode, execute, and writeback.

[00142] The CPU 405 can be part of a circuit, such as an integrated circuit. One or more other components of the system 401 can be included in the circuit. In some cases, the circuit is an application specific integrated circuit (ASIC).

[00143] The storage unit 415 can store files, such as drivers, libraries and saved programs. The storage unit 415 can store user data, *e.g.*, user preferences and user programs. The computer system 401 in some cases can include one or more additional data storage units that are external to the computer system 401, such as located on a remote server that is in communication with the computer system 401 through an intranet or the Internet.

[00144] The computer system 401 can communicate with one or more remote computer systems through the network 430. For instance, the computer system 401 can communicate with a remote computer system of a user (*e.g.*, a health care provider). Examples of remote computer systems include personal computers (*e.g.*, portable PC), slate or tablet PC's (*e.g.*, Apple® iPad, Samsung® Galaxy Tab), telephones, Smart phones (*e.g.*, Apple® iPhone, Android-enabled device, Blackberry®), or personal digital assistants. The user can access the computer system 401 via the network 430.

[00145] Methods as described herein can be implemented by way of machine (*e.g.*, computer processor) executable code stored on an electronic storage location of the computer system 401, such as, for example, on the memory 410 or electronic storage unit 415. The machine executable or machine-readable code can be provided in the form of software. During use, the code can be executed by the processor 405. In some cases, the code can be retrieved from the storage unit 415

and stored on the memory 410 for ready access by the processor 405. In some situations, the electronic storage unit 415 can be precluded, and machine-executable instructions are stored on memory 410.

[00146] The code can be pre-compiled and configured for use with a machine having a processor adapted to execute the code, or can be compiled during runtime. The code can be supplied in a programming language that can be selected to enable the code to execute in a pre-compiled or as-compiled fashion.

[00147] Aspects of the systems and methods provided herein, such as the computer system 401, can be embodied in programming. Various aspects of the technology may be thought of as “products” or “articles of manufacture” typically in the form of machine (or processor) executable code and/or associated data that is carried on or embodied in a type of machine readable medium. Machine-executable code can be stored on an electronic storage unit, such as memory (e.g., read-only memory, random-access memory, flash memory) or a hard disk. “Storage” type media can include any or all of the tangible memory of the computers, processors or the like, or associated modules thereof, such as various semiconductor memories, tape drives, disk drives and the like, which may provide non-transitory storage at any time for the software programming. All or portions of the software may at times be communicated through the Internet or various other telecommunication networks. Such communications, for example, may enable loading of the software from one computer or processor into another, for example, from a management server or host computer into the computer platform of an application server. Thus, another type of media that may bear the software elements includes optical, electrical, and electromagnetic waves, such as used across physical interfaces between local devices, through wired and optical landline networks and over various air-links. The physical elements that carry such waves, such as wired or wireless links, optical links or the like, also may be considered as media bearing the software. As used herein, unless restricted to non-transitory, tangible “storage” media, terms such as computer or machine “readable medium” refer to any medium that participates in providing instructions to a processor for execution.

[00148] Hence, a machine readable medium, such as computer-executable code, may take many forms, including but not limited to, a tangible storage medium, a carrier wave medium or physical transmission medium. Non-volatile storage media include, for example, optical or magnetic disks, such as any of the storage devices in any computer(s) or the like, such as may be used to implement the databases, *etc.* shown in the drawings. Volatile storage media include

dynamic memory, such as main memory of such a computer platform. Tangible transmission media include coaxial cables; copper wire and fiber optics, including the wires that comprise a bus within a computer system. Carrier-wave transmission media may take the form of electric or electromagnetic signals, or acoustic or light waves such as those generated during radio frequency (RF) and infrared (IR) data communications. Common forms of computer-readable media therefore include for example: a floppy disk, a flexible disk, hard disk, magnetic tape, any other magnetic medium, a CD-ROM, DVD or DVD-ROM, any other optical medium, punch cards paper tape, any other physical storage medium with patterns of holes, a RAM, a ROM, a PROM and EPROM, a FLASH-EPROM, any other memory chip or cartridge, a carrier wave transporting data or instructions, cables or links transporting such a carrier wave, or any other medium from which a computer may read programming code and/or data. Many of these forms of computer readable media may be involved in carrying one or more sequences of one or more instructions to a processor for execution.

[00149] The computer system 401 can include or be in communication with an electronic display 435 that comprises a user interface (UI) 440 for providing, for example, Raman image data, Raman spectral data, temporal Raman profiles, and models. Examples of UI's include, without limitation, a graphical user interface (GUI) and web-based user interface.

[00150] Methods and systems of the present disclosure can be implemented by way of one or more algorithms. An algorithm can be implemented by way of software upon execution by the central processing unit 405. The algorithm can, for example, obtain a Raman image of tooth samples, analyze Raman spectra spatially across tooth samples, generate a temporal Raman profile, process data using trained models, and predict a subject's diagnostic status with respect to a disease or disorder.

EXAMPLES

[00151] Example 1: Dynamic Raman spectroscopy profiles in tooth samples for determining autism spectrum disorder (ASD) disease risk

[00152] Using methods and systems of the present disclosure, dynamic Raman spectroscopy profiles in tooth samples were generated and subsequently analyzed to determine a disease risk in a subject. Generally, the temporal dynamics of biological response (*e.g.*, physiological responses) were found to be imprinted in samples (*e.g.*, tooth samples), and can be analyzed to determine

disease risk in a subject. Dynamic Raman spectroscopy profiles were generated during a time period that comprised fetal (prenatal) development and early childhood in two sets of children—a first set with autism spectrum disorder and a second set without autism spectrum disorder (ASD). The dynamic Raman spectroscopy profiles were analyzed to reveal novel features therein, which accurately distinguished the autism cases from controls. For example, early life spectroscopic signatures were found to reveal a disease risk of ASD in later life. As a comparison, a clinical diagnosis of autism is usually determined around the age of 3 to 4 years.

[00153] A primary tooth sample was obtained from each child subject. The tooth samples were sectioned open and Raman spectroscopy signals were measured on the tooth samples in order to develop temporal Raman spectroscopy profiles indicative of physiological response over the prenatal and postnatal period. The temporal profiles were analyzed using machine learning algorithms of the present disclosure to train highly accurate classifiers to determine disease risk (*e.g.*, autism).

[00154] **FIG. 5** shows an example of classifier accuracy of diagnosing autism spectrum disorder (ASD) utilizing features derived from application of RQA to ICA-derived dimensions of the Raman waveform, as indicated by an experimental Receiver Operating Characteristics (ROC) curve for evaluating accuracy of the disclosed method of evaluating a subject for autism spectrum disorder. A ROC curve can be used for evaluating a performance of a binary classifier. A ROC curve is plotted as sensitivity (also called as a true positive rate) against specificity (also called as a true negative rate). A perfect classifier may have a 100% sensitivity and 100% specificity and an Area-Under-the-Curve (AUC) of 1.0. As shown in **FIG. 5**, the classifier configured to determine the presence of ASD in a subject based on dynamic Raman RQA dynamic profile had an Area-Under-the-Curve (AUC) of the receiver operating characteristic (ROC) of 0.861, with a 95% confidence interval (CI) of 0.769 to 0.954. The receiver operating characteristic (ROC) shows how sensitivity and specificity values of the classifier change as varying thresholds are assigned to probabilistic projections.

[00155] Therefore, analysis of Raman spectroscopic signatures using methods and systems of the present disclosure successfully determined the disease risk of autism with greater than 0.86 AUC, using only spectroscopic signatures measured on non-invasively obtained biological samples (*e.g.*, tooth samples) from child subjects. These results demonstrate that dynamics of human physiology in early life are linked to disease later on, which can be accurately detected and profiled using methods and systems of the present disclosure.

[00156] Example 2: Dynamic Raman spectroscopy profiles in tooth samples for determining Amyotrophic Lateral Sclerosis disease risk

[00157] Using methods and systems of the present disclosure, dynamic Raman spectroscopy profiles in tooth samples were generated and subsequently analyzed to determine a disease risk in a subject. Generally, the temporal dynamics of biological response (*e.g.*, physiological responses) were found to be imprinted in samples (*e.g.*, tooth samples), and can be analyzed to determine disease risk in a subject. Dynamic Raman spectroscopy profiles were generated during a time period that comprised early childhood and adolescence in two sets of adults—a first set with amyotrophic lateral sclerosis (ALS) and a second set without ALS. The dynamic Raman spectroscopy profiles were analyzed to reveal novel features therein, which accurately distinguished the ALS cases from controls. For example, early life spectroscopic signatures were found to reveal a disease risk of ALS in later life.

[00158] A permanent tooth sample was obtained from each adult subject. The tooth samples were sectioned open and Raman spectroscopy signals were measured on the tooth samples in order to develop temporal Raman spectroscopy profiles indicative of physiological response over the early childhood and adolescence period. The temporal profiles were analyzed using machine learning algorithms of the present disclosure to train highly accurate classifiers to determine disease risk (*e.g.*, ALS).

[00159] **FIG. 6** shows an example of classifier accuracy of diagnosing ALS utilizing features derived from application of RQA to ICA-derived dimensions of the Raman waveform, as indicated by an experimental Receiver Operating Characteristics (ROC) curve for evaluating accuracy of the disclosed method of evaluating a subject for autism spectrum disorder. A ROC curve can be used for evaluating a performance of a binary classifier. A ROC curve is plotted as sensitivity (also called as a true positive rate) against specificity (also called as a true negative rate). A perfect classifier may have a 100% sensitivity and 100% specificity and an Area-Under-the-Curve (AUC) of 1.0. As shown in **FIG. 6**, the classifier configured to determine the presence of ASD in a subject based on dynamic Raman RQA dynamic profile had an Area-Under-the-Curve (AUC) of the receiver operating characteristic (ROC) of 0.880, with a 95% confidence interval (CI) of 0.658 to 1.000. The receiver operating characteristic (ROC) shows how sensitivity and specificity values of the classifier change as varying thresholds are assigned to probabilistic projections.

[00160] Therefore, analysis of Raman spectroscopic signatures using methods and systems of the present disclosure successfully determined the disease risk of ALS with greater than 0.88 AUC, using only spectroscopic signatures measured on biological samples (*e.g.*, tooth samples) from adults. These results demonstrate that dynamics of human physiology in early life are linked to disease later on, which can be accurately detected and profiled using methods and systems of the present disclosure.

[00161] Although the methods described elsewhere herein, show steps or sets of operations in accordance with embodiments, a person of ordinary skill in the art will recognize many variations based on the teaching described herein. The steps may be completed in a different order. Steps may be added or omitted. Some of the steps may comprise sub-steps. Many of the steps may be repeated as often as beneficial.

[00162] One or more of the steps of each of the methods or sets of operations may be performed with circuitry as described herein, for example, one or more of the processor or logic circuitry such as programmable array logic for a field programmable gate array. The circuitry may be programmed to provide one or more of the steps of each of the methods or sets of operations, and the program may comprise program instructions stored on a computer readable memory or programmed steps of the logic circuitry such as the programmable array logic or the field programmable gate array, for example.

[00163] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. It is not intended that the invention be limited by the specific examples provided within the specification. While the invention has been described with reference to the aforementioned specification, the descriptions and illustrations of the embodiments herein are not meant to be construed in a limiting sense. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. Furthermore, it shall be understood that all aspects of the invention are not limited to the specific depictions, configurations or relative proportions set forth herein which depend upon a variety of conditions and variables. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is therefore contemplated that the invention shall also cover any such alternatives, modifications, variations, or equivalents. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

EMBODIMENTS

[00164] Embodiment 1. A method for predicting a subject's diagnostic status with respect to a disease or disorder, comprising: (a) exposing a biological sample of a subject to a light source; (b) acquiring a plurality of Raman spectra from the biological sample; (c) processing the plurality of Raman spectra to generate a spatial map of the plurality of Raman spectra; and (d) predicting the subject's diagnostic status with respect to the disease or disorder based at least in part on the spatial map of the plurality of Raman spectra.

[00165] Embodiment 2. The method of embodiment 1, wherein the biological sample comprises a tooth sample, a hair sample, a nail sample, or any combination thereof.

[00166] Embodiment 3. The method of embodiment 1 or 2, further comprising detecting or monitoring changes in a temporal stress profile of the spatial map that are indicative of a temporal response of the subject.

[00167] Embodiment 4. The method of embodiment 3, wherein the temporal response comprises a biological response, a physiological response, an anatomical response, a treatment response, a stress-related response, or a combination thereof response.

[00168] Embodiment 5. The method of any one of embodiments 1-4, wherein the plurality of Raman spectra comprises from about 200 to about 3700 wave numbers.

[00169] Embodiment 6. The method of any one of embodiments 1-5, wherein acquiring comprises using a Raman spectroscopy microscope.

[00170] Embodiment 7. The method of embodiment 6, wherein the Raman spectroscopy microscope comprises an 50X air coupled objective, 63X water immersion coupled objection, or any combination thereof.

[00171] Embodiment 8. The method of any one of embodiments 1-7, wherein the light source comprises a laser, wherein the laser comprises a wavelength of about 785 nm, a wavelength of about 532 nm, or any combination thereof.

[00172] Embodiment 9. The method of any one of embodiments 1-8, wherein the acquiring is performed using an integration time of about 0.2 seconds to about 0.3 seconds.

[00173] Embodiment 10. The method of any one of embodiments 1-9, wherein the acquiring comprises moving the biological sample with a step size of about 2 microns to about 5 microns, subsequent to acquiring a Raman spectrum of the plurality of Raman spectra.

[00174] Embodiment 11. The method of any one of embodiments 1-10, wherein the disease or disorder comprises autism spectrum disorder (ASD), attention deficit/hyperactivity disorder (ADHD), amyotrophic lateral sclerosis (ALS), schizophrenia, irritable bowel disease (IBD), pediatric kidney disease, kidney transplant rejection, pediatric cancer or any combination thereof.

[00175] Embodiment 12. The method of any one of embodiments 1-10, wherein the disease or disorder comprises the ASD.

[00176] Embodiment 13. The method of any one of embodiments 1-12, wherein predicting the subject's diagnostic status with respect to the disease or disorder comprises processing the spatial map using a trained model.

[00177] Embodiment 14. The method of embodiment 13, wherein the trained model is selected from the group consisting of: a neural network algorithm, a support vector machine algorithm, a decision tree algorithm, an unsupervised clustering algorithm, a supervised clustering algorithm, a regression algorithm, a gradient-boosting algorithm, and any combination thereof.

[00178] Embodiment 15. The method of embodiment 13, wherein the trained model comprises a gradient-boosted ensemble model.

[00179] Embodiment 16. The method of embodiment 13, wherein the trained model is configured to process one or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, determination of a linear slope of the temporal stress profile, determination of a plurality of non-linear parameters describing curvature of the temporal stress profile, determination of an abrupt change in intensity of the temporal stress profile, determination of one or more changes in a baseline intensity of the temporal stress profile, determination of a change of a frequency-domain representation of the temporal stress profile, determination of a change of the power-spectral domain representation of the temporal stress profile, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, determination of a maximum Lyapunov exponent and any combination thereof.

[00180] Embodiment 17. The method of embodiment 3, wherein the trained model is configured to process two or more features selected from the group consisting of laminarity,

entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, determination of a linear slope of the temporal stress profile, determination of a plurality of non-linear parameters describing curvature of the temporal stress profile, determination of an abrupt change in intensity of the temporal stress profile, determination of one or more changes in a baseline intensity of the temporal stress profile, determination of a change of a frequency-domain representation of the temporal stress profile, determination of a change of the power-spectral domain representation of the temporal stress profile, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, determination of a maximum Lyapunov exponent and any combination thereof.

[00181] Embodiment 18. The method of any one of embodiments 1-17, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a sensitivity of at least about 80%.

[00182] Embodiment 19. The method of embodiment 1, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a specificity of at least about 80%.

[00183] Embodiment 20. The method of embodiment 1, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a positive predictive value of at least about 80%.

[00184] Embodiment 21. The method of embodiment 1, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a negative predictive value of at least about 80%.

[00185] Embodiment 22. The method of embodiment 1, wherein the trained model predicts diagnostic status with respect to the disease or disorder with an Area Under the Receiver Operating Characteristic (AUROC) of at least about 0.80.

[00186] Embodiment 23. A device comprising one or more processors, and memory storing one or more programs for execution by the one or more processors, the one or more programs comprising instructions for: (a) sampling each respective position in a plurality of positions along a reference line on a biological sample of a subject associated with a Raman signature of the subject, thereby obtaining a plurality of Raman spectra, each Raman spectrum in the plurality of Raman spectra corresponding to a different position in the plurality of positions, and each

position in the plurality of positions representing a different period of growth of the biological sample associated with the Raman signature; (b) analyzing each of the plurality of Raman spectra across a reference line on the biological sample thereby obtaining a first dataset; (c) deriving a respective second dataset from the corresponding plurality of the Raman spectra measurements, each respective feature in the corresponding set of features being determined by a sequential variation in the Raman spectra; and (d) processing the features using a trained machine learning model to predict a subject's diagnostic status with respect to disease or disorder associated with the Raman signature.

[00187] Embodiment 24. The device of embodiment 23, wherein the biological sample comprises a tooth sample, a hair sample, a nail sample, or any combination thereof.

[00188] Embodiment 25. The device of embodiment 23 or 24, wherein the instructions further comprise detecting or monitoring changes in the Raman spectra across the plurality of positions indicative of a temporal response of the subject.

[00189] Embodiment 26. The device of embodiment 25, wherein the temporal response comprises a biological response, a physiological response, an anatomical response, a treatment response, a stress-related response, or a combination thereof response.

[00190] Embodiment 27. The device of any one of embodiments 23-26, wherein the plurality of Raman spectra comprises from about 200 to about 3700 wave numbers.

[00191] Embodiment 28. The device of any one of embodiments 23-27, wherein sampling comprises using a Raman spectroscopy microscope.

[00192] Embodiment 29. The device of embodiment 28, wherein the Raman spectroscopy microscope comprises an 50X air coupled objective, 63X water immersion coupled objection, or any combination thereof.

[00193] Embodiment 30. The device of embodiment 23, wherein the sampling comprises exposing the biological sample to a light source to generate the Raman spectra of the plurality of Raman spectra at the plurality of positions.

[00194] Embodiment 31. The device of embodiment 30, wherein the light source comprises a laser, wherein the laser comprises a wavelength of about 785 nm, a wavelength of about 532 nm, or any combination thereof.

[00195] Embodiment 32. The device of any one of embodiments 23-31, wherein the instructions further comprise translating, wherein translating comprises moving the biological sample with a step size of about 2 microns to about 5 microns from a first position to a second

position of the plurality of positions subsequent to acquiring a Raman spectrum of the plurality of Raman spectra.

[00196] Embodiment 33. The device of embodiment 32, wherein the translating is performed using an integration time of about 0.2 seconds to about 0.3 seconds.

[00197] Embodiment 34. The device of any one of embodiments 23-33, wherein the disease or disorder comprises autism spectrum disorder (ASD), attention deficit/hyperactivity disorder (ADHD), amyotrophic lateral sclerosis (ALS), schizophrenia, irritable bowel disease (IBD), pediatric kidney disease, kidney transplant rejection, pediatric cancer or any combination thereof.

[00198] Embodiment 35. The device of any one of embodiments 23-33, wherein the disease or disorder comprises autism spectrum disorder (ASD).

[00199] Embodiment 36. The device of any one of embodiments 23-35, wherein predicting a subject's diagnostic status with respect to the disease or disorder comprises processing changes in the Raman spectra across the plurality of positions with a trained model.

[00200] Embodiment 37. The device of embodiment 36, wherein the trained model is selected from the group consisting of: a neural network algorithm, a support vector machine algorithm, a decision tree algorithm, an unsupervised clustering algorithm, a supervised clustering algorithm, a regression algorithm, a gradient-boosting algorithm, and any combination thereof.

[00201] Embodiment 38. The device of embodiment 36, wherein the trained model comprises a gradient-boosted ensemble model.

[00202] Embodiment 39. The device of embodiment 36, wherein the trained model is configured to process one or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, determination of a linear slope of the plurality of Raman spectra across a reference line, determination of a plurality of non-linear parameters describing curvature of the plurality of Raman spectra across a reference line, determination of an abrupt change in intensity of the plurality of Raman spectra across a reference line, determination of one or more changes in a baseline intensity of the plurality of Raman spectra across a reference line, determination of a change of a frequency-domain representation of the plurality of Raman spectra across a reference line, determination of a change of the power-spectral domain representation of the plurality of Raman spectra across a reference line, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters,

determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, determination of a maximum Lyapunov exponent and any combination thereof.

[00203] Embodiment 40. The device of embodiment 36, wherein the trained model is configured to process two or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, determination of a linear slope of the plurality of Raman spectra across a reference line, determination of a plurality of non-linear parameters describing curvature of the plurality of Raman spectra across a reference line, determination of an abrupt change in intensity of the plurality of Raman spectra across a reference line, determination of one or more changes in a baseline intensity of the plurality of Raman spectra across a reference line, determination of a change of a frequency-domain representation of the plurality of Raman spectra across a reference line, determination of a change of the power-spectral domain representation of the plurality of Raman spectra across a reference line, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, determination of a maximum Lyapunov exponent and any combination thereof.

[00204] Embodiment 41. The device of embodiment 23, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a sensitivity of at least about 80%.

[00205] Embodiment 42. The device of embodiment 23, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a specificity of at least about 80%.

[00206] Embodiment 43. The device of embodiment 23, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a positive predictive value of at least about 80%.

[00207] Embodiment 44. The device of embodiment 23, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a negative predictive value of at least about 80%.

[00208] Embodiment 45. The device of embodiment 23, wherein the trained model predicts diagnostic status with respect to the disease or disorder with an Area Under the Receiver Operating Characteristic (AUROC) of at least about 0.80.

[00209] Embodiment 46. A non-transitory computer readable storage medium and one or more computer programs embedded therein for classification, the one or more computer programs comprising instructions which, when executed by a computer system, cause the computer system to perform a method comprising: (a) sampling each respective position in a plurality of positions along a reference line on a biological sample of a subject associated with a Raman signature of the subject, thereby obtaining a plurality of Raman spectra, each Raman spectra in the plurality of Raman spectra corresponding to a different position in the plurality of positions, and each position in the plurality of positions representing a different period of growth of the biological sample associated with the Raman signature; (b) analyzing each of the plurality of Raman spectra across a reference line on the biological sample thereby obtaining a first dataset; (c) deriving a respective second dataset from the corresponding plurality of the Raman spectra measurements, each respective feature in the corresponding set of features being determined by a sequential variation in the Raman spectra; and (d) processing the features using a trained model to predict a subject's diagnostic status with respect to disease or disorder associated with the Raman signature.

[00210] Embodiment 47. The non-transitory computer readable storage medium of embodiment 46, wherein the biological sample comprises a tooth sample, a hair sample, a nail sample, or any combination thereof.

[00211] Embodiment 48. The non-transitory computer readable storage medium of embodiment 46 or 47, wherein the method further comprise detecting or monitoring changes in the Raman spectra across the plurality of positions indicative of a temporal response of the subject.

[00212] Embodiment 49. The non-transitory computer readable storage medium of embodiment 48, wherein the temporal response comprises a biological response, a physiological response, an anatomical response, a treatment response, a stress-related response, or a combination thereof response.

[00213] Embodiment 50. The non-transitory computer readable storage medium of any one of embodiments 46-49, wherein the plurality of Raman spectra comprises from about 200 to about 3700 wave numbers.

[00214] Embodiment 51. The non-transitory computer readable storage medium of any one of embodiments 46-50, wherein sampling comprises using a Raman spectroscopy microscope.

[00215] Embodiment 52. The non-transitory computer readable storage medium of embodiment 51, wherein the Raman spectroscopy microscope comprises an 50X air coupled objective, 63X water immersion coupled objection, or any combination thereof.

[00216] Embodiment 53. The non-transitory computer readable storage medium of any one of embodiments 46-52, wherein sampling comprises exposing the biological sample to a light source to generate the Raman spectra of the plurality of Raman spectra at the plurality of positions.

[00217] Embodiment 54. The non-transitory computer readable storage medium of embodiment 53, wherein the light source comprises a laser, wherein the laser comprises a wavelength of about 785 nm, a wavelength of about 532 nm, or any combination thereof.

[00218] Embodiment 55. The non-transitory computer readable storage medium of any one of embodiments 46-54, wherein the instructions further comprise translating, wherein translating comprises moving the biological sample with a step size of about 2 microns to about 5 microns from a first position to a second position of the plurality of positions subsequent to acquiring a Raman spectrum of the plurality of Raman spectra.

[00219] Embodiment 56. The non-transitory computer readable storage medium of embodiment 55, wherein translating is performed using an integration time of about 0.2 seconds to about 0.3 seconds.

[00220] Embodiment 57. The non-transitory computer readable storage medium of any one of embodiments 46-56, wherein the disease or disorder comprises autism spectrum disorder (ASD), attention deficit/hyperactivity disorder (ADHD), amyotrophic lateral sclerosis (ALS), schizophrenia, irritable bowel disease (IBD), pediatric kidney disease, kidney transplant rejection, pediatric cancer or any combination thereof.

[00221] Embodiment 58. The non-transitory computer readable storage medium of any one of embodiments 46-56, wherein the disease or disorder comprises autism spectrum disorder (ASD).

[00222] Embodiment 59. The non-transitory computer readable storage medium of any one of embodiments 46-58, wherein predicting a subject's diagnostic status with respect to the disease or disorder comprises processing changes in the Raman spectra across the plurality of positions with a trained model.

[00223] Embodiment 60. The non-transitory computer readable storage medium of embodiment 59, wherein the trained model is selected from the group consisting of: a neural network algorithm, a support vector machine algorithm, a decision tree algorithm, an

unsupervised clustering algorithm, a supervised clustering algorithm, a regression algorithm, a gradient-boosting algorithm, and any combination thereof.

[00224] Embodiment 61. The non-transitory computer readable storage medium of embodiment 59, wherein the trained model comprises a gradient-boosted ensemble model.

[00225] Embodiment 62. The non-transitory computer readable storage medium of embodiment 59, wherein the trained model is configured to process one or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, determination of a linear slope of the plurality of Raman spectra across a reference line, determination of a plurality of non-linear parameters describing curvature of the plurality of Raman spectra across a reference line, determination of an abrupt change in intensity of the plurality of Raman spectra across a reference line, determination of one or more changes in a baseline intensity of the plurality of Raman spectra across a reference line, determination of a change of a frequency-domain representation of the plurality of Raman spectra across a reference line, determination of a change of the power-spectral domain representation of the plurality of Raman spectra across a reference line, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, determination of a maximum Lyapunov exponent and any combination thereof.

[00226] Embodiment 63. The non-transitory computer readable storage medium of embodiment 59, wherein the trained model is configured to process two or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, determination of a linear slope of the plurality of Raman spectra across a reference line, determination of a plurality of non-linear parameters describing curvature of the plurality of Raman spectra across a reference line, determination of an abrupt change in intensity of the plurality of Raman spectra across a reference line, determination of one or more changes in a baseline intensity of the plurality of Raman spectra across a reference line, determination of a change of a frequency-domain representation of the plurality of Raman spectra across a reference line, determination of a change of the power-spectral domain representation of the plurality of Raman spectra across a reference line, determination of one or more recurrence quantification analysis parameters,

determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, determination of a maximum Lyapunov exponent and any combination thereof.

[00227] Embodiment 64. The non-transitory computer readable storage medium of embodiment 46, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a sensitivity of at least about 80%.

[00228] Embodiment 65. The non-transitory computer readable storage medium of embodiment 46, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a specificity of at least about 80%.

[00229] Embodiment 66. The non-transitory computer readable storage medium of embodiment 46, wherein the instruction further comprise predicting a subject's diagnostic status with respect to the disease or disorder with a positive predictive value of at least about 80%.

[00230] Embodiment 67. The non-transitory computer readable storage medium of embodiment 46, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a positive predictive value of at least about 80%.

[00231] Embodiment 68. The non-transitory computer readable storage medium of embodiment 46, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a negative predictive value of at least about 80%.

[00232] Embodiment 69. A method for training a model, comprising: at a computer system having one or more processors, and memory storing one or more programs for execution by the one or more processors: (a) for each respective training subject in a plurality of training subjects, wherein a first subset of training subjects in the plurality of training subjects have a first diagnostic status corresponding to having a first biological condition associated with a Raman signature and a second subset of training subjects in the plurality of training subjects have a second diagnostic status corresponding to not having the first biological condition associated with the Raman signature: (i) sampling each respective position in a plurality of positions along a reference line on a biological sample of the subject associated with the Raman signature of the subject, thereby obtaining a plurality of Raman spectra, each Raman spectra in the plurality of Raman spectra corresponding to a different position in the plurality of positions, and each position in the plurality of positions represent a different period of growth of the biological sample of the subject associated with the Raman signature; (ii) analyzing each Raman spectra

across a reference line on biological sample thereby obtaining a first dataset; and (iii) deriving a respective second dataset from the corresponding plurality of Raman spectra, each respective feature in the corresponding set of features being determined by a sequential variation in Raman spectra; and (b) training an untrained or partially untrained model with (i) the corresponding set of features of each respective second dataset of each training subject in the plurality of training subjects and (ii) the corresponding diagnostic status of each training subject in the plurality of training subjects, selected from among the first diagnostic status and the second diagnostic status, thereby obtaining a trained model that provides an indication as to whether a test subject has the first biological condition associated with the Raman signature based on values for features in a set of features acquired from a biological sample associated with the Raman signature of the test subject.

[00233] Embodiment 70. The method of embodiment 69, wherein the trained model is selected from the group consisting of: a neural network algorithm, a support vector machine algorithm, a decision tree algorithm, an unsupervised clustering algorithm, a supervised clustering algorithm, a regression algorithm, a gradient-boosting algorithm, and any combination thereof.

[00234] Embodiment 71. The method of embodiment 69, wherein the trained model is multinomial classifier.

[00235] Embodiment 72. The method of embodiment 69, wherein the trained model is a binomial classifier.

[00236] Embodiment 73. The method of embodiment 69, wherein the first biological condition is selected from the group consisting of autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD), amyotrophic lateral sclerosis (ALS), schizophrenia, irritable bowel disease (IBD), pediatric kidney disease, kidney transplant rejection, and pediatric cancer.

[00237] Embodiment 74. The method of any one of embodiments 69-73, wherein evaluating the test subject for the first biological condition associated with a Raman signature further includes discriminating between the first biological condition associated with the Raman signature and a second biological condition associated with the Raman signature distinct from the first biological condition associated with the Raman signature.

- [00238] Embodiment 75. The method of embodiment 74, wherein the first biological condition is autism spectrum disorder and the second biological condition is attention-deficit/hyperactivity disorder.
- [00239] Embodiment 76. The method of any one of embodiments 69-75, wherein the test subject is a human.
- [00240] Embodiment 77. The method of embodiment 76, wherein the human is less than 12 years old.
- [00241] Embodiment 78. The method of embodiment 76, wherein the human is less than 1 year old.
- [00242] Embodiment 79. The method of any one of embodiments 69-78, wherein the corresponding biological sample associated with the Raman signature of the respective training subject is selected from the group consisting of a hair shaft, a tooth, and a nail.
- [00243] Embodiment 80. The method of embodiment 79, wherein the corresponding biological sample associated with the Raman signature of the respective training subject is the hair shaft, and wherein the reference line corresponds to a longitudinal direction of the hair shaft.
- [00244] Embodiment 81. The method of embodiment 79, wherein the corresponding biological sample associated with the Raman signature of the respective training subject is the tooth, and wherein the reference line corresponds to a direction across the growth bands, including the neonatal line of the tooth.
- [00245] Embodiment 82. The method of any one of embodiments 69-81, wherein the corresponding plurality of positions is sequenced such that a first position in the corresponding plurality of positions along the corresponding biological sample of the respective training subject corresponds to a position closest to a tip of the corresponding biological sample of the respective training subject.
- [00246] Embodiment 83. The method of any one of embodiments 69-82, wherein each trace in the corresponding plurality of Raman spectral measurements includes a plurality of data points, each data point being an instance of the respective position in the plurality of positions.
- [00247] Embodiment 84. The method of any one of embodiments 69-83, wherein the corresponding set of features is selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax.

[00248] Embodiment 85. The method of any one of embodiments 69-83, wherein the corresponding plurality of positions includes at least 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, or 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, 10000, or more than 10000 positions.

[00249] Embodiment 86. The method of any one of embodiments 69-85, wherein the trained model is configured to process one or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, determination of a linear slope of the plurality of Raman spectra across a reference line, determination of a plurality of non-linear parameters describing curvature of the plurality of Raman spectra across a reference line, determination of an abrupt change in intensity of the plurality of Raman spectra across a reference line, determination of one or more changes in a baseline intensity of the plurality of Raman spectra across a reference line, determination of a change of a frequency-domain representation of the plurality of Raman spectra across a reference line, determination of a change of the power-spectral domain representation of the plurality of Raman spectra across a reference line, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, determination of a maximum Lyapunov exponent and any combination thereof.

[00250] Embodiment 87. The method of any one of embodiments 69-85, wherein the trained model is configured to process two or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, determination of a linear slope of the plurality of Raman spectra across a reference line, determination of a plurality of non-linear parameters describing curvature of the plurality of Raman spectra across a reference line, determination of an abrupt change in intensity of the plurality of Raman spectra across a reference line, determination of one or more changes in a baseline intensity of the plurality of Raman spectra across a reference line, determination of a change of a frequency-domain representation of the plurality of Raman spectra across a reference line, determination of a change of the power-spectral domain representation of the plurality of Raman spectra across a reference line, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-

recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, determination of a maximum Lyapunov exponent and any combination thereof.

WHAT IS CLAIMED IS:

1. A method for predicting a subject's diagnostic status with respect to a disease or disorder, comprising:
 - (a) exposing a biological sample of a subject to a light source;
 - (b) acquiring a plurality of Raman spectra from the biological sample;
 - (c) processing the plurality of Raman spectra to generate a spatial map of the plurality of Raman spectra; and
 - (d) predicting the subject's diagnostic status with respect to the disease or disorder based at least in part on the spatial map of the plurality of Raman spectra.
2. The method of claim 1, wherein the biological sample comprises a tooth sample, a hair sample, a nail sample, or any combination thereof.
3. The method of claim 1 or 2, further comprising detecting or monitoring changes in a temporal stress profile of the spatial map that are indicative of a temporal response of the subject.
4. The method of claim 3, wherein the temporal response comprises a biological response, a physiological response, an anatomical response, a treatment response, a stress-related response, or a combination thereof response.
5. The method of any one of claims 1-4, wherein the plurality of Raman spectra comprises from about 200 to about 3700 wave numbers.
6. The method of any one of claims 1-5, wherein acquiring comprises using a Raman spectroscopy microscope.
7. The method of claim 6, wherein the Raman spectroscopy microscope comprises an 50X air coupled objective, 63X water immersion coupled objection, or any combination thereof.

8. The method of any one of claims 1-7, wherein the light source comprises a laser, wherein the laser comprises a wavelength of about 785 nm, a wavelength of about 532 nm, or any combination thereof.
9. The method of any one of claims 1-8, wherein the acquiring is performed using an integration time of about 0.2 seconds to about 0.3 seconds.
10. The method of any one of claims 1-9, wherein the acquiring comprises moving the biological sample with a step size of about 2 microns to about 5 microns, subsequent to acquiring a Raman spectrum of the plurality of Raman spectra.
11. The method of any one of claims 1-10, wherein the disease or disorder comprises autism spectrum disorder (ASD), attention deficit/hyperactivity disorder (ADHD), amyotrophic lateral sclerosis (ALS), schizophrenia, irritable bowel disease (IBD), pediatric kidney disease, kidney transplant rejection, pediatric cancer or any combination thereof.
12. The method of any one of claims 1-10, wherein the disease or disorder comprises the ASD.
13. The method of any one of claims 1-12, wherein predicting the subject's diagnostic status with respect to the disease or disorder comprises processing the spatial map using a trained model.
14. The method of claim 13, wherein the trained model is selected from the group consisting of: a neural network algorithm, a support vector machine algorithm, a decision tree algorithm, an unsupervised clustering algorithm, a supervised clustering algorithm, a regression algorithm, a gradient-boosting algorithm, and any combination thereof.
15. The method of claim 13, wherein the trained model comprises a gradient-boosted ensemble model.
16. The method of claim 3, wherein the trained model is configured to process one or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean

diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, determination of a linear slope of the temporal stress profile, determination of a plurality of non-linear parameters describing curvature of the temporal stress profile, determination of an abrupt change in intensity of the temporal stress profile, determination of one or more changes in a baseline intensity of the temporal stress profile, determination of a change of a frequency-domain representation of the temporal stress profile, determination of a change of the power-spectral domain representation of the temporal stress profile, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, determination of a maximum Lyapunov exponent and any combination thereof.

17. The method of claim 16, wherein the trained model is configured to process two or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, determination of a linear slope of the temporal stress profile, determination of a plurality of non-linear parameters describing curvature of the temporal stress profile, determination of an abrupt change in intensity of the temporal stress profile, determination of one or more changes in a baseline intensity of the temporal stress profile, determination of a change of a frequency-domain representation of the temporal stress profile, determination of a change of the power-spectral domain representation of the temporal stress profile, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, determination of a maximum Lyapunov exponent and any combination thereof.

18. The method of any one of claims 1-17, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a sensitivity of at least about 80%.

19. The method of claim 1, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a specificity of at least about 80%.
20. The method of claim 1, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a positive predictive value of at least about 80%.
21. The method of claim 1, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a negative predictive value of at least about 80%.
22. The method of claim 1, wherein the trained model predicts diagnostic status with respect to the disease or disorder with an Area Under the Receiver Operating Characteristic (AUROC) of at least about 0.80.
23. A device comprising one or more processors, and memory storing one or more programs for execution by the one or more processors, the one or more programs comprising instructions for:
- (a) sampling each respective position in a plurality of positions along a reference line on a biological sample of a subject associated with a Raman signature of the subject, thereby obtaining a plurality of Raman spectra, each Raman spectrum in the plurality of Raman spectra corresponding to a different position in the plurality of positions, and each position in the plurality of positions representing a different period of growth of the biological sample associated with the Raman signature;
 - (b) analyzing each of the plurality of Raman spectra across a reference line on the biological sample thereby obtaining a first dataset;
 - (c) deriving a respective second dataset from the corresponding plurality of the Raman spectra measurements, each respective feature in the corresponding set of features being determined by a sequential variation in the Raman spectra; and
 - (d) processing the features using a trained model to predict a subject's diagnostic status with respect to disease or disorder associated with the Raman signature.
24. The device of claim 23, wherein the biological sample comprises a tooth sample, a hair sample, a nail sample, or any combination thereof.

25. The device of claim 23 or 24, wherein the instructions further comprise detecting or monitoring changes in the Raman spectra across the plurality of positions indicative of a temporal response of the subject.
26. The device of claim 25, wherein the temporal response comprises a biological response, a physiological response, an anatomical response, a treatment response, a stress-related response, or a combination thereof response.
27. The device of any one of claims 23-26, wherein the plurality of Raman spectra comprises from about 200 to about 3700 wave numbers.
28. The device of any one of claims 23-27, wherein sampling comprises using a Raman spectroscopy microscope.
29. The device of claim 28, wherein the Raman spectroscopy microscope comprises an 50X air coupled objective, 63X water immersion coupled objection, or any combination thereof.
30. The device of claim 23, wherein the sampling comprises exposing the biological sample to a light source to generate the Raman spectra of the plurality of Raman spectra at the plurality of positions.
31. The device of claim 30, wherein the light source comprises a laser, wherein the laser comprises a wavelength of about 785 nm, a wavelength of about 532 nm, or any combination thereof.
32. The device of any one of claims 23-31, wherein the instructions further comprise translating, wherein translating comprises moving the biological sample with a step size of about 2 microns to about 5 microns from a first position to a second position of the plurality of positions subsequent to acquiring a Raman spectrum of the plurality of Raman spectra.

33. The device of claim 32, wherein the translating is performed using an integration time of about 0.2 seconds to about 0.3 seconds.

34. The device of any one of claims 23-33, wherein the disease or disorder comprises autism spectrum disorder (ASD), attention deficit/hyperactivity disorder (ADHD), amyotrophic lateral sclerosis (ALS), schizophrenia, irritable bowel disease (IBD), pediatric kidney disease, kidney transplant rejection, pediatric cancer or any combination thereof.

35. The device of any one of claims 23-33, wherein the disease or disorder comprises autism spectrum disorder (ASD).

36. The device of any one of claims 23-35, wherein predicting a subject's diagnostic status with respect to the disease or disorder comprises processing changes in the Raman spectra across the plurality of positions with a trained model.

37. The device of claim 36, wherein the trained model is selected from the group consisting of: a neural network algorithm, a support vector machine algorithm, a decision tree algorithm, an unsupervised clustering algorithm, a supervised clustering algorithm, a regression algorithm, a gradient-boosting algorithm, and any combination thereof.

38. The device of claim 36, wherein the trained model comprises a gradient-boosted ensemble model.

39. The device of claim 36, wherein the trained model is configured to process one or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, determination of a linear slope of the plurality of Raman spectra across a reference line, determination of a plurality of non-linear parameters describing curvature of the plurality of Raman spectra across a reference line, determination of an abrupt change in intensity of the plurality of Raman spectra across a reference line, determination of one or more changes in a baseline intensity of the plurality of Raman spectra across a reference line, determination of a change of a frequency-domain representation of the plurality of Raman spectra across a reference line, determination of

a change of the power-spectral domain representation of the plurality of Raman spectra across a reference line, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, determination of a maximum Lyapunov exponent and any combination thereof.

40. The device of claim 36, wherein the trained model is configured to process two or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, determination of a linear slope of the plurality of Raman spectra across a reference line, determination of a plurality of non-linear parameters describing curvature of the plurality of Raman spectra across a reference line, determination of an abrupt change in intensity of the plurality of Raman spectra across a reference line, determination of one or more changes in a baseline intensity of the plurality of Raman spectra across a reference line, determination of a change of a frequency-domain representation of the plurality of Raman spectra across a reference line, determination of a change of the power-spectral domain representation of the plurality of Raman spectra across a reference line, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, determination of a maximum Lyapunov exponent and any combination thereof.

41. The device of claim 23, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a sensitivity of at least about 80%.

42. The device of claim 23, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a specificity of at least about 80%.

43. The device of claim 23, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a positive predictive value of at least about 80%.

44. The device of claim 23, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a negative predictive value of at least about 80%.

45. The device of claim 23, wherein the trained model predicts diagnostic status with respect to the disease or disorder with an Area Under the Receiver Operating Characteristic (AUROC) of at least about 0.80.

46. A non-transitory computer readable storage medium and one or more computer programs embedded therein for classification, the one or more computer programs comprising instructions which, when executed by a computer system, cause the computer system to perform a method comprising:

(a) sampling each respective position in a plurality of positions along a reference line on a biological sample of a subject associated with a Raman signature of the subject, thereby obtaining a plurality of Raman spectra, each Raman spectra in the plurality of Raman spectra corresponding to a different position in the plurality of positions, and each position in the plurality of positions representing a different period of growth of the biological sample associated with the Raman signature;

(b) analyzing each of the plurality of Raman spectra across a reference line on the biological sample thereby obtaining a first dataset;

(c) deriving a respective second dataset from the corresponding plurality of the Raman spectra measurements, each respective feature in the corresponding set of features being determined by a sequential variation in the Raman spectra; and

(d) processing the features using a trained model to predict a subject's diagnostic status with respect to disease or disorder associated with the Raman signature.

47. The non-transitory computer readable storage medium of claim 46, wherein the biological sample comprises a tooth sample, a hair sample, a nail sample, or any combination thereof.

48. The non-transitory computer readable storage medium of claim 46 or 47, wherein the method further comprise detecting or monitoring changes in the Raman spectra across the plurality of positions indicative of a temporal response of the subject.

49. The non-transitory computer readable storage medium of claim 48, wherein the temporal response comprises a biological response, a physiological response, an anatomical response, a treatment response, a stress-related response, or a combination thereof response.

50. The non-transitory computer readable storage medium of any one of claims 46-49, wherein the plurality of Raman spectra comprises from about 200 to about 3700 wave numbers.

51. The non-transitory computer readable storage medium of any one of claims 46-50, wherein sampling comprises using a Raman spectroscopy microscope.

52. The non-transitory computer readable storage medium of claim 51, wherein the Raman spectroscopy microscope comprises an 50X air coupled objective, 63X water immersion coupled objection, or any combination thereof.

53. The non-transitory computer readable storage medium of any one of claims 46-52, wherein sampling comprises exposing the biological sample to a light source to generate the Raman spectra of the plurality of Raman spectra at the plurality of positions.

54. The non-transitory computer readable storage medium of claim 53, wherein the light source comprises a laser, wherein the laser comprises a wavelength of about 785 nm, a wavelength of about 532 nm, or any combination thereof.

55. The non-transitory computer readable storage medium of any one of claims 46-54, wherein the instructions further comprise translating, wherein translating comprises moving the biological sample with a step size of about 2 microns to about 5 microns from a first position to a second position of the plurality of positions subsequent to acquiring a Raman spectrum of the plurality of Raman spectra.

56. The non-transitory computer readable storage medium of claim 55, wherein translating is performed using an integration time of about 0.2 seconds to about 0.3 seconds.

57. The non-transitory computer readable storage medium of any one of claims 46-56, wherein the disease or disorder comprises autism spectrum disorder (ASD), attention deficit/hyperactivity disorder (ADHD), amyotrophic lateral sclerosis (ALS), schizophrenia, irritable bowel disease (IBD), pediatric kidney disease, kidney transplant rejection, pediatric cancer or any combination thereof.

58. The non-transitory computer readable storage medium of any one of claims 46-56, wherein the disease or disorder comprises autism spectrum disorder (ASD).

59. The non-transitory computer readable storage medium of any one of claims 46-58, wherein predicting a subject's diagnostic status with respect to the disease or disorder comprises processing changes in the Raman spectra across the plurality of positions with a trained model.

60. The non-transitory computer readable storage medium of claim 59, wherein the trained model is selected from the group consisting of: a neural network algorithm, a support vector machine algorithm, a decision tree algorithm, an unsupervised clustering algorithm, a supervised clustering algorithm, a regression algorithm, a gradient-boosting algorithm, and any combination thereof.

61. The non-transitory computer readable storage medium of claim 59, wherein the trained model comprises a gradient-boosted ensemble model.

62. The non-transitory computer readable storage medium of claim 59, wherein the trained model is configured to process one or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, determination of a linear slope of the plurality of Raman spectra across a reference line, determination of a plurality of non-linear parameters describing curvature of the plurality of Raman spectra across a reference line, determination of an abrupt change in intensity of the plurality of Raman spectra across a reference line, determination of one or more changes in a baseline intensity of the plurality of Raman spectra across a reference line, determination of a change of a frequency-domain representation of the plurality of Raman spectra across a reference line, determination of a change of the power-spectral domain

representation of the plurality of Raman spectra across a reference line, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, determination of a maximum Lyapunov exponent and any combination thereof.

63. The non-transitory computer readable storage medium of claim 59, wherein the trained model is configured to process two or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, determination of a linear slope of the plurality of Raman spectra across a reference line, determination of a plurality of non-linear parameters describing curvature of the plurality of Raman spectra across a reference line, determination of an abrupt change in intensity of the plurality of Raman spectra across a reference line, determination of one or more changes in a baseline intensity of the plurality of Raman spectra across a reference line, determination of a change of a frequency-domain representation of the plurality of Raman spectra across a reference line, determination of a change of the power-spectral domain representation of the plurality of Raman spectra across a reference line, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, determination of a maximum Lyapunov exponent and any combination thereof.

64. The non-transitory computer readable storage medium of claim 46, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a sensitivity of at least about 80%.

65. The non-transitory computer readable storage medium of claim 46, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a specificity of at least about 80%.

66. The non-transitory computer readable storage medium of claim 46, wherein the instruction further comprise predicting a subject's diagnostic status with respect to the disease or disorder with a positive predictive value of at least about 80%.

67. The non-transitory computer readable storage medium of claim 46, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a positive predictive value of at least about 80%.

68. The non-transitory computer readable storage medium of claim 46, wherein the trained model predicts diagnostic status with respect to the disease or disorder with a negative predictive value of at least about 80%.

69. A method for training a model, comprising:

at a computer system having one or more processors, and memory storing one or more programs for execution by the one or more processors:

(a) for each respective training subject in a plurality of training subjects, wherein a first subset of training subjects in the plurality of training subjects have a first diagnostic status corresponding to having a first biological condition associated with a Raman signature and a second subset of training subjects in the plurality of training subjects have a second diagnostic status corresponding to not having the first biological condition associated with the Raman signature:

(i) sampling each respective position in a plurality of positions along a reference line on a biological sample of the subject associated with the Raman signature of the subject, thereby obtaining a plurality of Raman spectra, each Raman spectra in the plurality of Raman spectra corresponding to a different position in the plurality of positions, and each position in the plurality of positions represent a different period of growth of the biological sample of the subject associated with the Raman signature;

(ii) analyzing each Raman spectra across a reference line on biological sample thereby obtaining a first dataset; and

(iii) deriving a respective second dataset from the corresponding plurality of Raman spectra, each respective feature in the corresponding set of features being determined by a sequential variation in Raman spectra; and

(b) training an untrained or partially untrained model with (i) the corresponding set of features of each respective second dataset of each training subject in the plurality of training subjects and (ii) the corresponding diagnostic status of each training subject in the plurality of training subjects, selected from among the first diagnostic status and the second diagnostic status, thereby obtaining a trained model that provides an indication as to whether a test subject has the first biological condition associated with the Raman signature based on values for features in a set of features acquired from a biological sample associated with the Raman signature of the test subject.

70. The method of claim 69, wherein the trained model is selected from the group consisting of: a neural network algorithm, a support vector machine algorithm, a decision tree algorithm, an unsupervised clustering algorithm, a supervised clustering algorithm, a regression algorithm, a gradient-boosting algorithm, and any combination thereof.

71. The method of claim 69, wherein the trained model is multinomial classifier.

72. The method of claim 69, wherein the trained model is a binomial classifier.

73. The method of claim 69, wherein the first biological condition is selected from the group consisting of autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD), amyotrophic lateral sclerosis (ALS), schizophrenia, irritable bowel disease (IBD), pediatric kidney disease, kidney transplant rejection, and pediatric cancer.

74. The method of any one of claims 69-73, wherein evaluating the test subject for the first biological condition associated with a Raman signature further includes discriminating between the first biological condition associated with the Raman signature and a second biological condition associated with the Raman signature distinct from the first biological condition associated with the Raman signature.

75. The method of claim 74, wherein the first biological condition is autism spectrum disorder and the second biological condition is attention-deficit/hyperactivity disorder.

76. The method of any one of claims 69-75, wherein the test subject is a human.
77. The method of claim 76, wherein the human is less than 12 years old.
78. The method of claim 76, wherein the human is less than 1 year old.
79. The method of any one of claims 69-78, wherein the corresponding biological sample associated with the Raman signature of the respective training subject is selected from the group consisting of a hair shaft, a tooth, and a nail.
80. The method of claim 79, wherein the corresponding biological sample associated with the Raman signature of the respective training subject is the hair shaft, and wherein the reference line corresponds to a longitudinal direction of the hair shaft.
81. The method of claim 79, wherein the corresponding biological sample associated with the Raman signature of the respective training subject is the tooth, and wherein the reference line corresponds to a direction across the growth bands, including the neonatal line of the tooth.
82. The method of any one of claims 69-81, wherein the corresponding plurality of positions is sequenced such that a first position in the corresponding plurality of positions along the corresponding biological sample of the respective training subject corresponds to a position closest to a tip of the corresponding biological sample of the respective training subject.
83. The method of any one of claims 69-82, wherein each trace in the corresponding plurality of Raman spectral measurements includes a plurality of data points, each data point being an instance of the respective position in the plurality of positions.
84. The method of any one of claims 69-83, wherein the corresponding set of features is selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax.

85. The method of any one of claims 69-83, wherein the corresponding plurality of positions includes at least 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, or 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, 10000, or more than 10000 positions.

86. The method of any one of claims 69-85, wherein the trained model is configured to process one or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, determination of a linear slope of the plurality of Raman spectra across a reference line, determination of a plurality of non-linear parameters describing curvature of the plurality of Raman spectra across a reference line, determination of an abrupt change in intensity of the plurality of Raman spectra across a reference line, determination of one or more changes in a baseline intensity of the plurality of Raman spectra across a reference line, determination of a change of a frequency-domain representation of the plurality of Raman spectra across a reference line, determination of a change of the power-spectral domain representation of the plurality of Raman spectra across a reference line, determination of one or more recurrence quantification analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, determination of a maximum Lyapunov exponent and any combination thereof.

87. The method of any one of claims 69-85, wherein the trained model is configured to process two or more features selected from the group consisting of laminarity, entropy, trapping time (TT), mean diagonal length (MDL), recurrence time (RT), Vmax, determinism, Lmax, determination of a linear slope of the plurality of Raman spectra across a reference line, determination of a plurality of non-linear parameters describing curvature of the plurality of Raman spectra across a reference line, determination of an abrupt change in intensity of the plurality of Raman spectra across a reference line, determination of one or more changes in a baseline intensity of the plurality of Raman spectra across a reference line, determination of a change of a frequency-domain representation of the plurality of Raman spectra across a reference line, determination of a change of the power-spectral domain representation of the plurality of Raman spectra across a reference line, determination of one or more recurrence quantification

analysis parameters, determination of one or more cross-recurrence quantification analysis parameters, determination of one or more joint recurrence quantification analysis parameters, determination of one or more multi-dimensional recurrence quantification analysis parameters, estimation of a Lyapunov spectra, determination of a maximum Lyapunov exponent and any combination thereof.