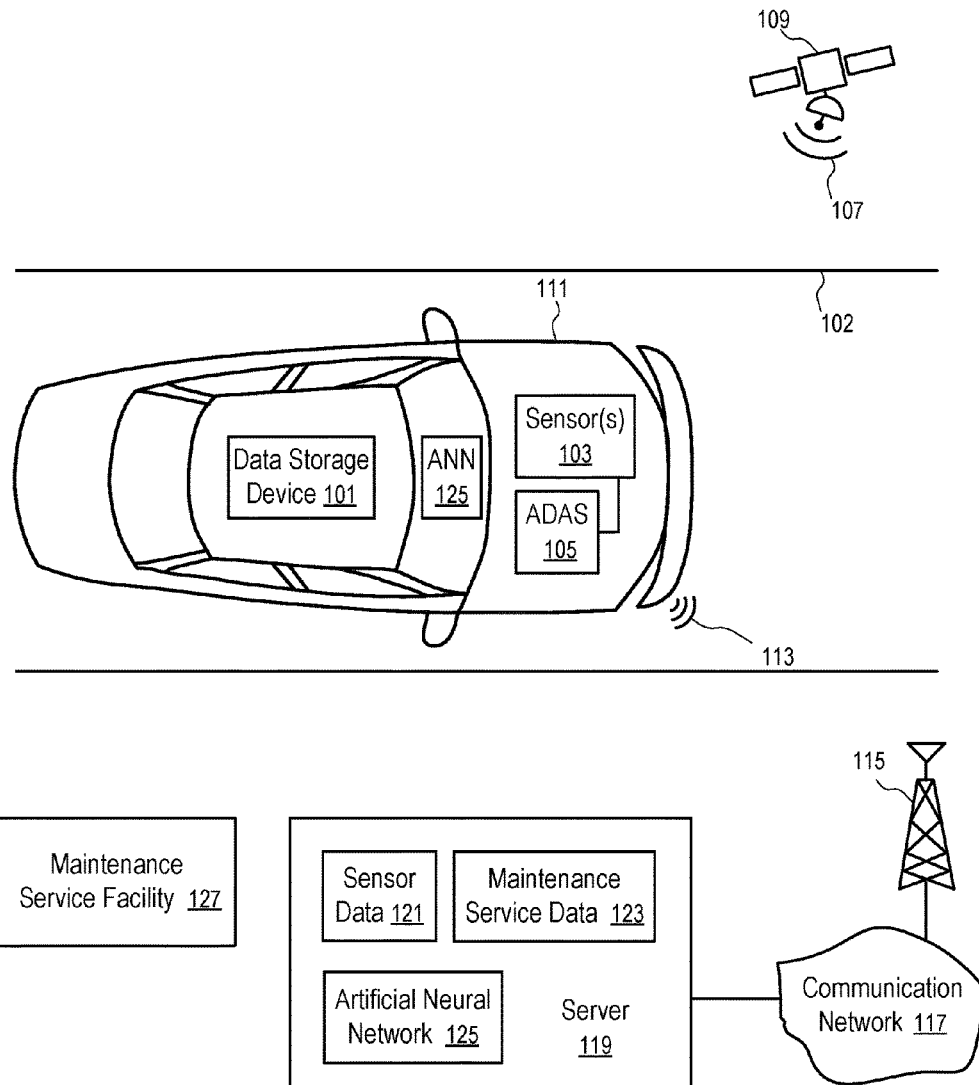




US 20210049480A1

(19) **United States**(12) **Patent Application Publication** (10) **Pub. No.: US 2021/0049480 A1**
(43) **Pub. Date: Feb. 18, 2021**(54) **PREDICTIVE MAINTENANCE OF
AUTOMOTIVE BATTERY**(71) Applicant: **Micron Technology, Inc.**, Boise, ID
(US)(72) Inventors: **Poorna Kale**, Folsom, CA (US);
Robert Richard Noel Bielby,
Placerville, CA (US)(21) Appl. No.: **16/538,097**(22) Filed: **Aug. 12, 2019****Publication Classification**(51) **Int. Cl.**
G06N 5/04 (2006.01)
G06N 3/04 (2006.01)
G06N 3/08 (2006.01)
G10L 25/51 (2006.01)
G10L 25/30 (2006.01)(52) **U.S. Cl.**CPC **G06N 5/04** (2013.01); **G06N 3/049**
(2013.01); **G10L 25/30** (2013.01); **G10L 25/51**
(2013.01); **G06N 3/08** (2013.01)(57) **ABSTRACT**

Systems, methods and apparatus of predictive maintenance of automotive battery. For example, a vehicle has: an electric motor; battery configured to power at least the electric motor; one or more sensors configured to measure operating parameters of the battery; an artificial neural network configured to analyze the operating parameters of the battery as a function of time to generate a result; and at least one processor configured to generate a suggestion for a maintenance service of the battery based on the result from the artificial neural network analyzing the operating parameters of the battery. For example, the electric motor can be part of an engine starter for an internal combustion engine, or a motor for an electric vehicle.



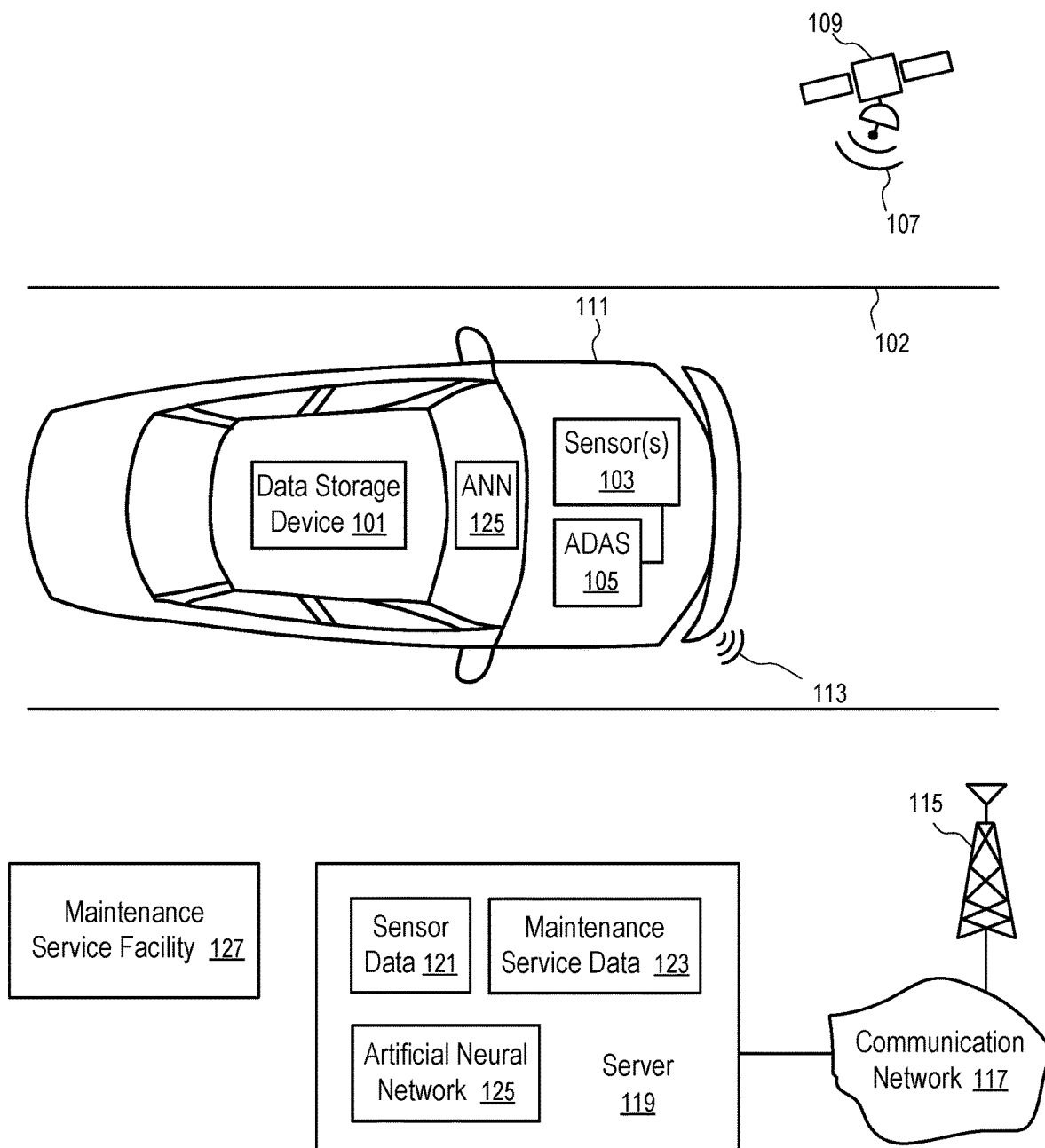


FIG. 1

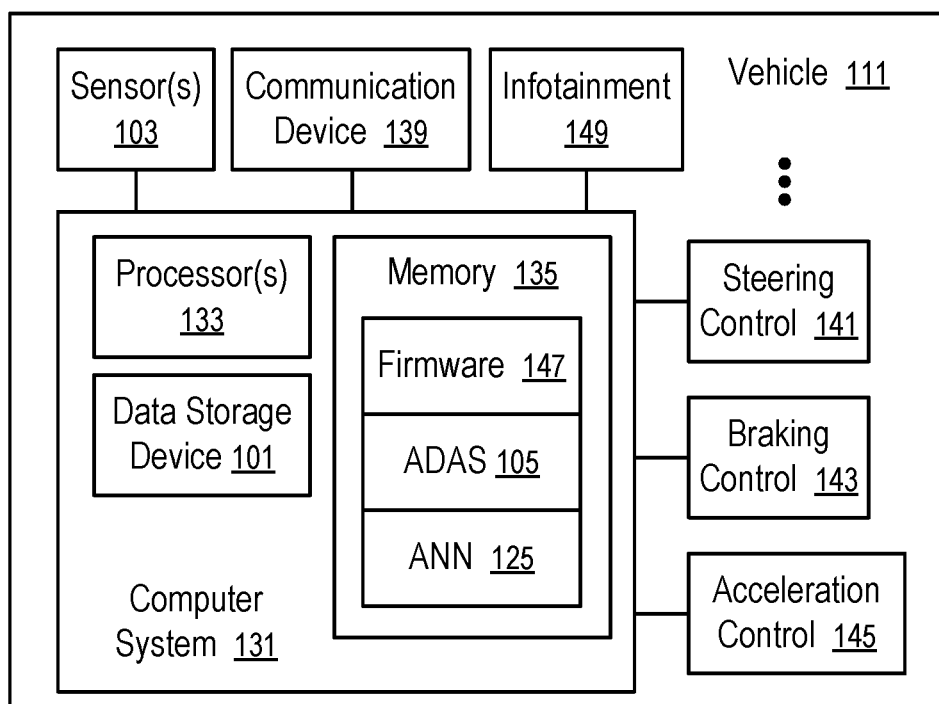


FIG. 2

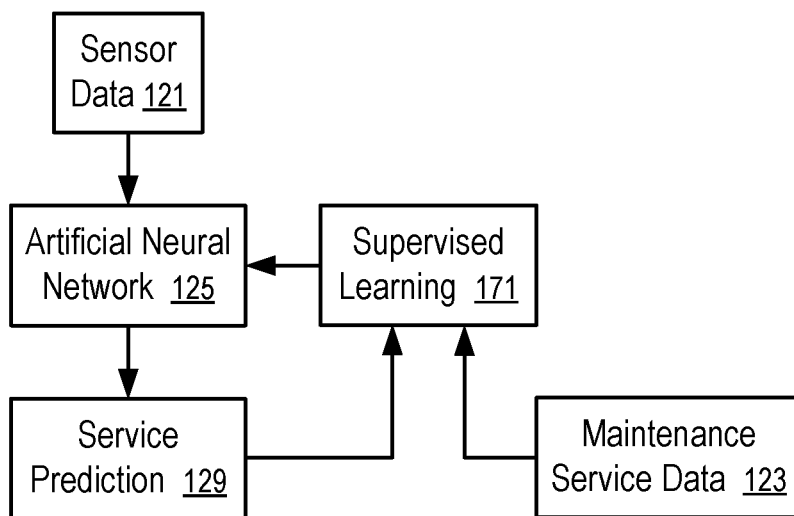


FIG. 3

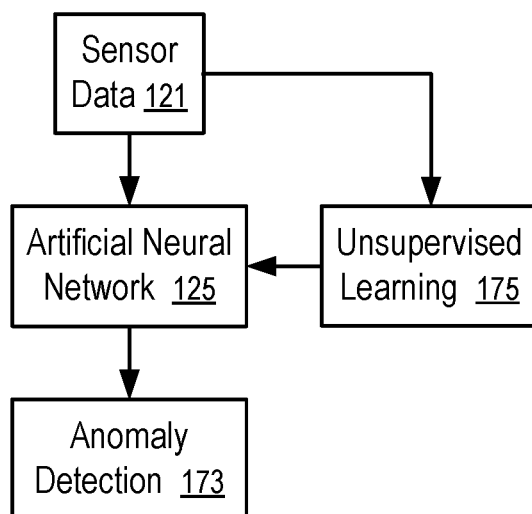


FIG. 4

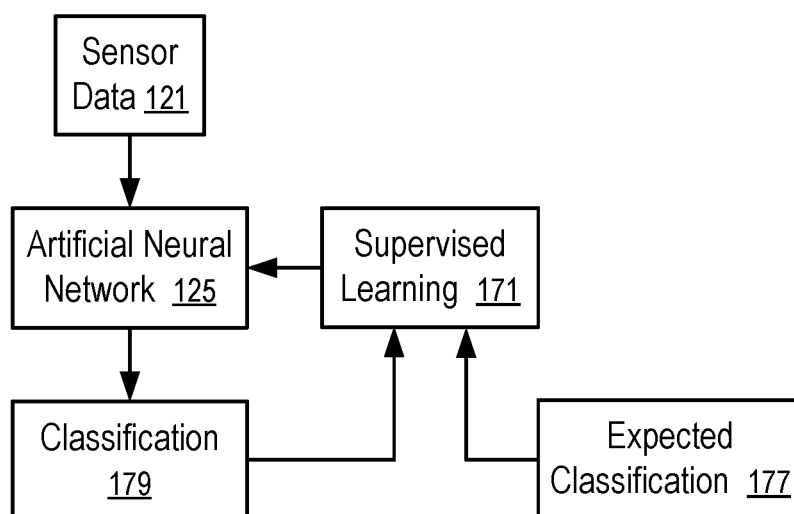


FIG. 5

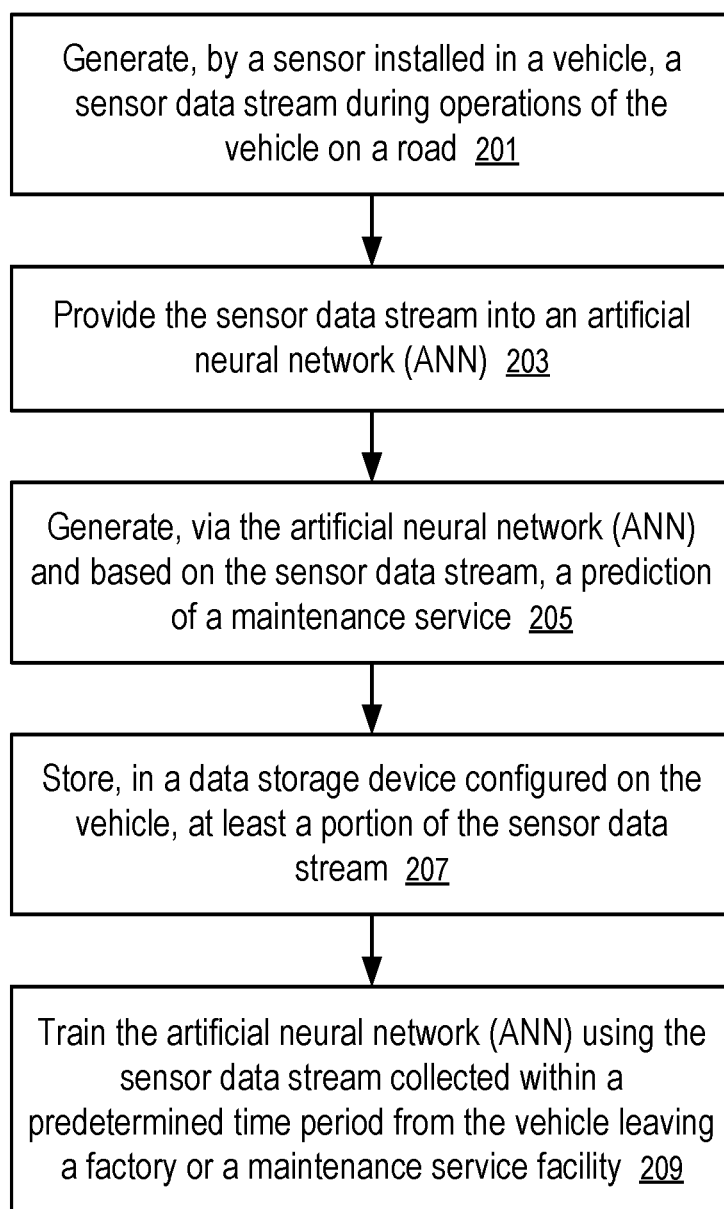


FIG. 6

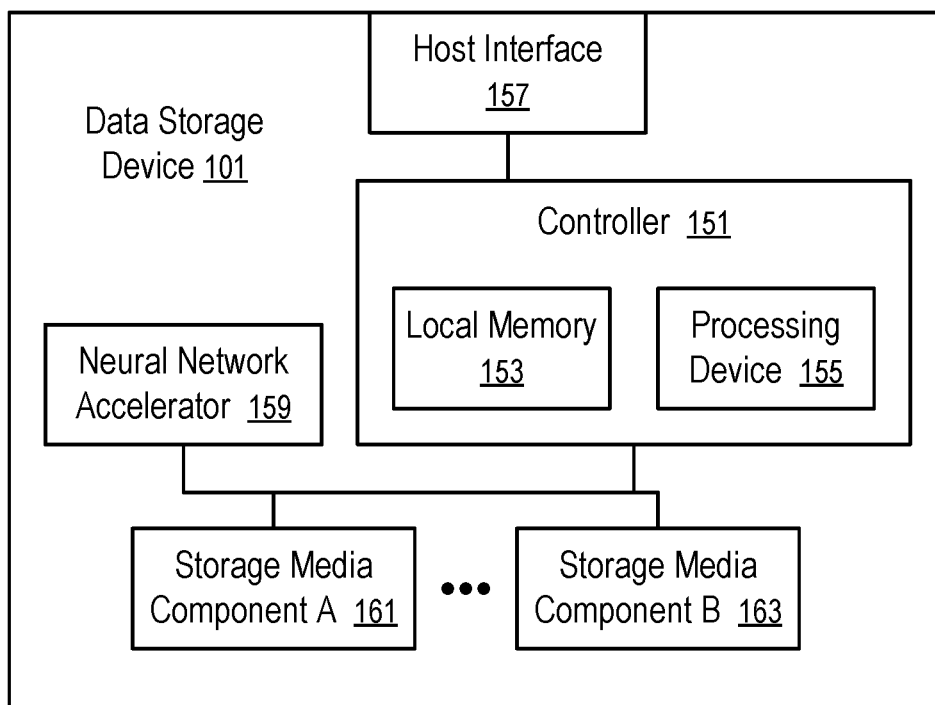


FIG. 7

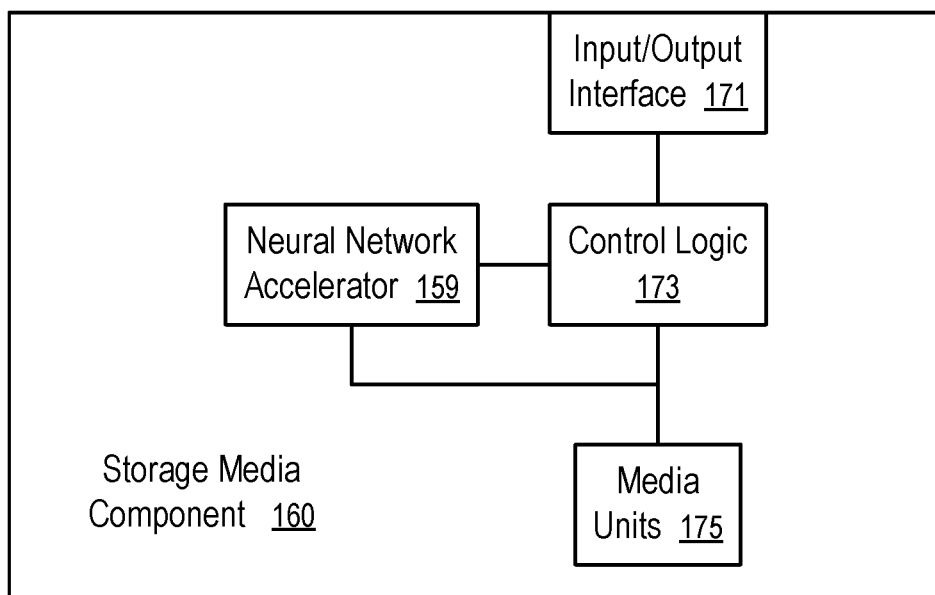


FIG. 8

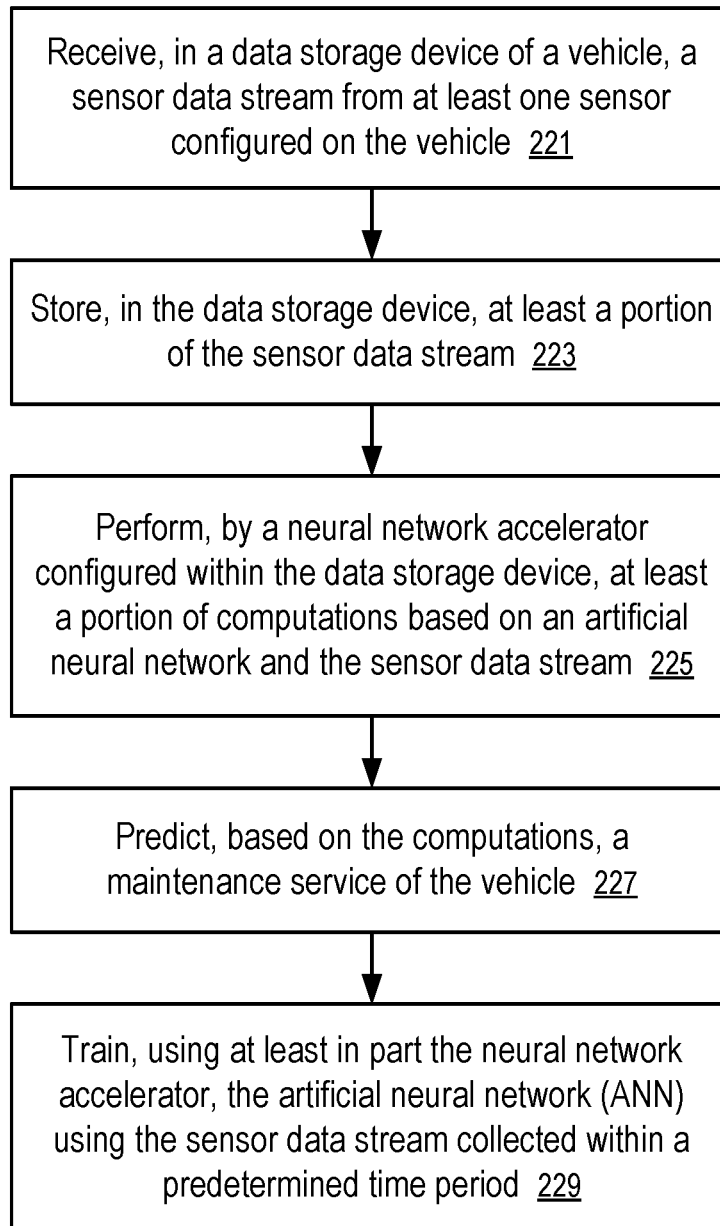


FIG. 9

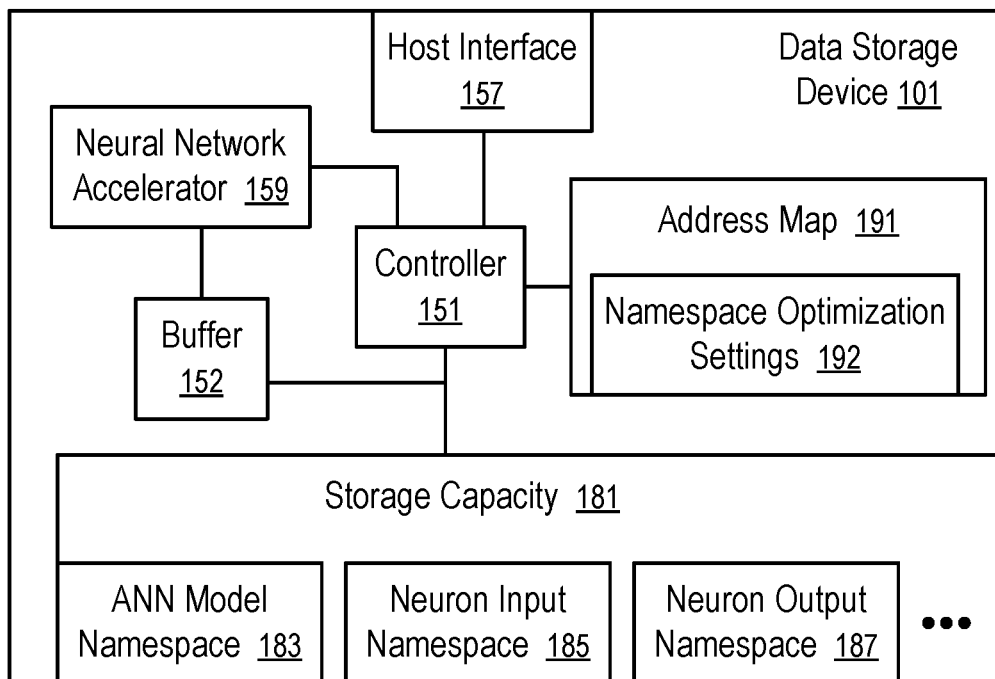


FIG. 10

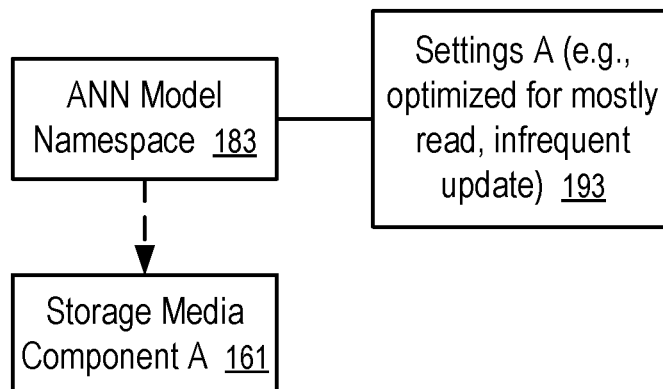


FIG. 11

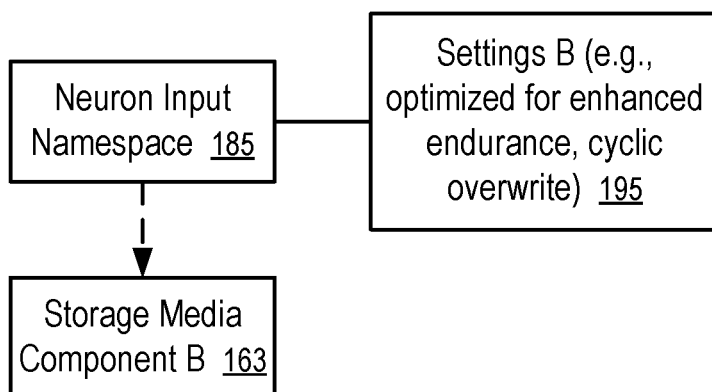


FIG. 12

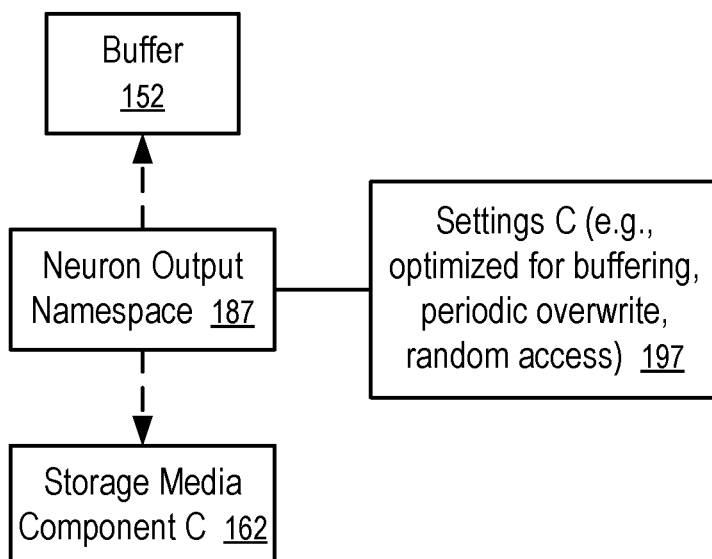


FIG. 13

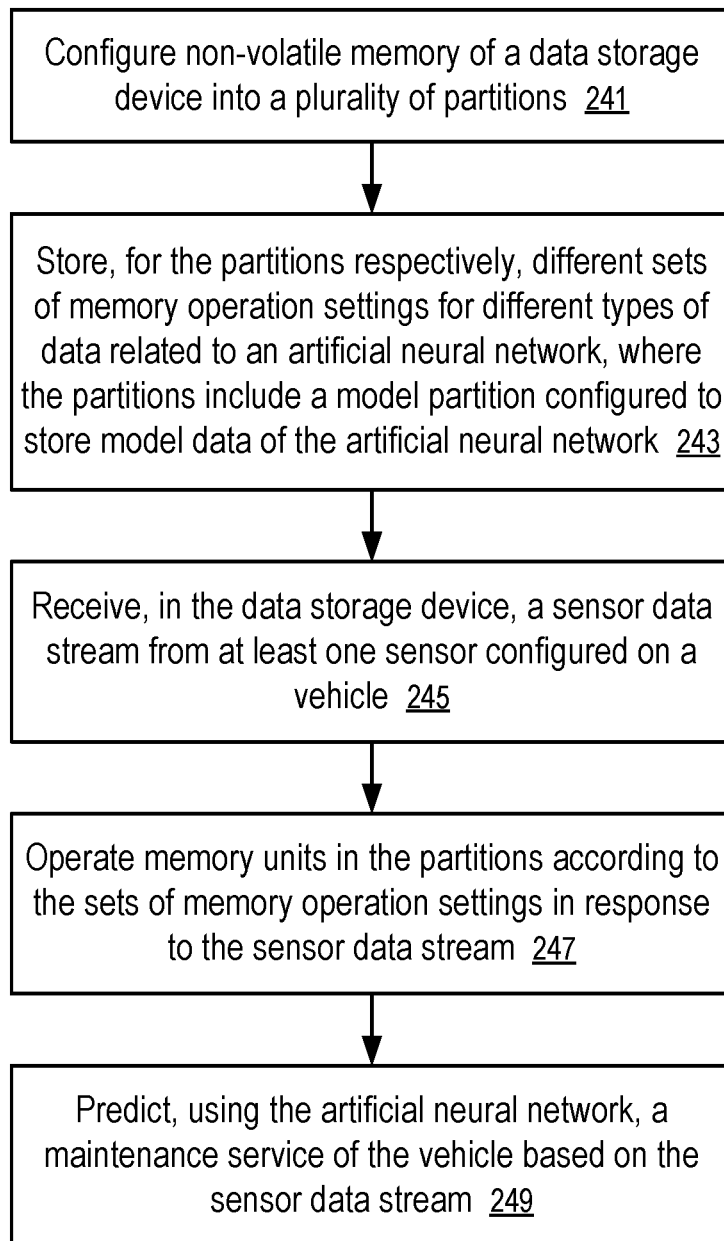


FIG. 14

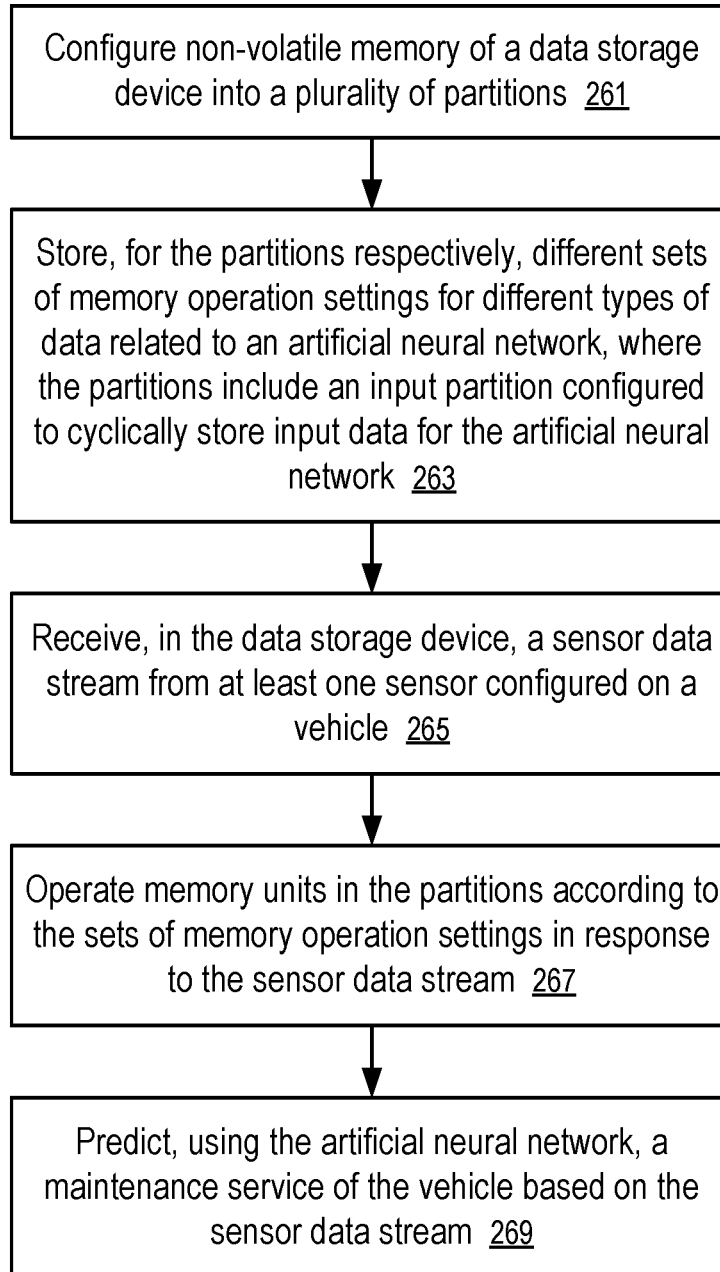


FIG. 15

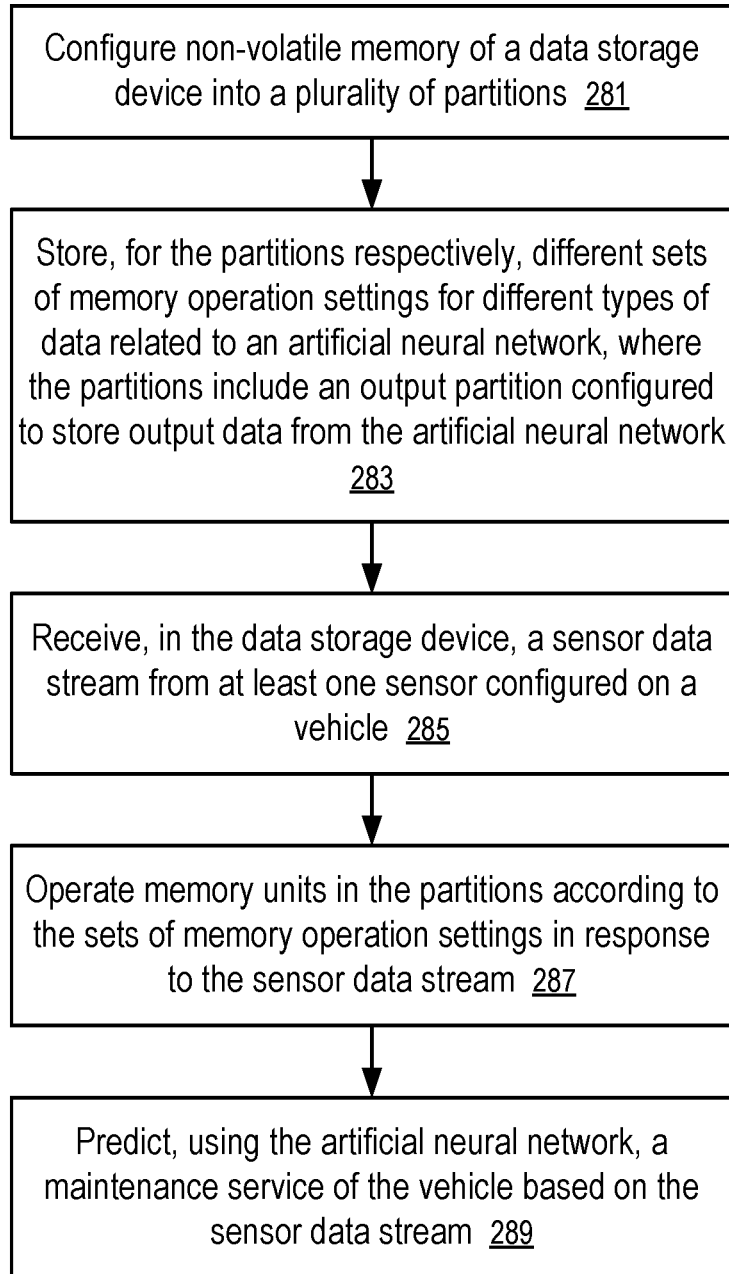


FIG. 16

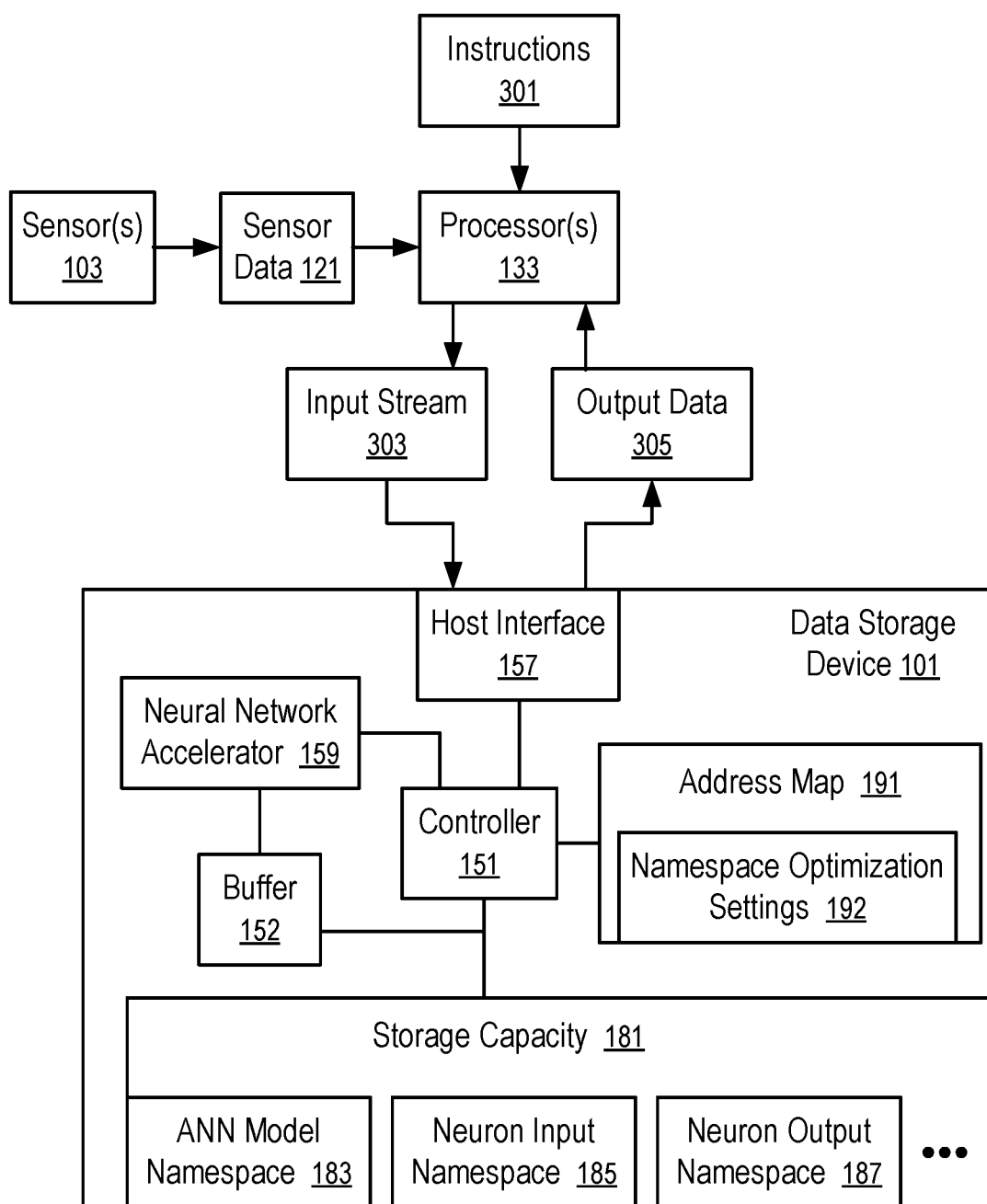


FIG. 17

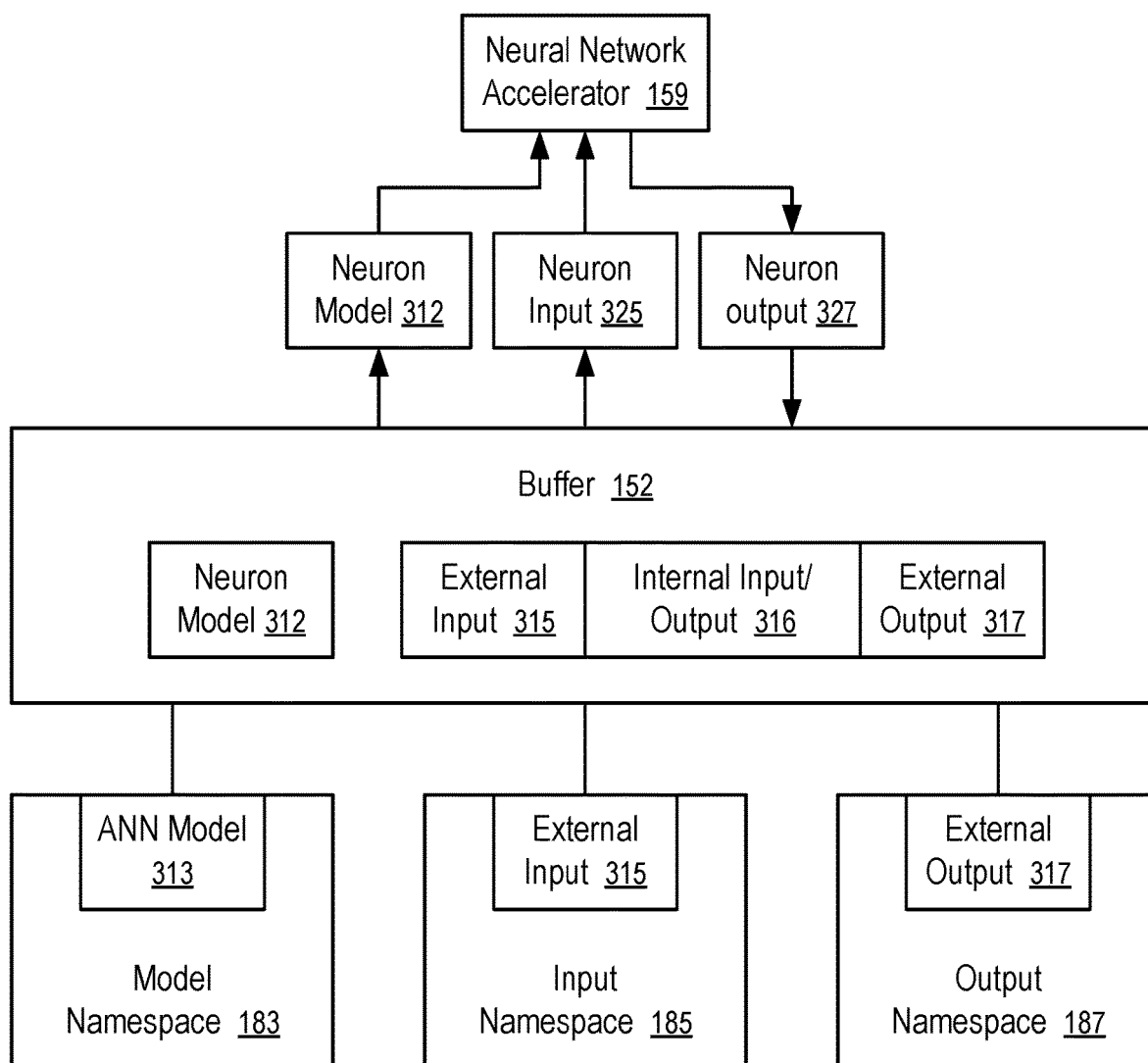


FIG. 18

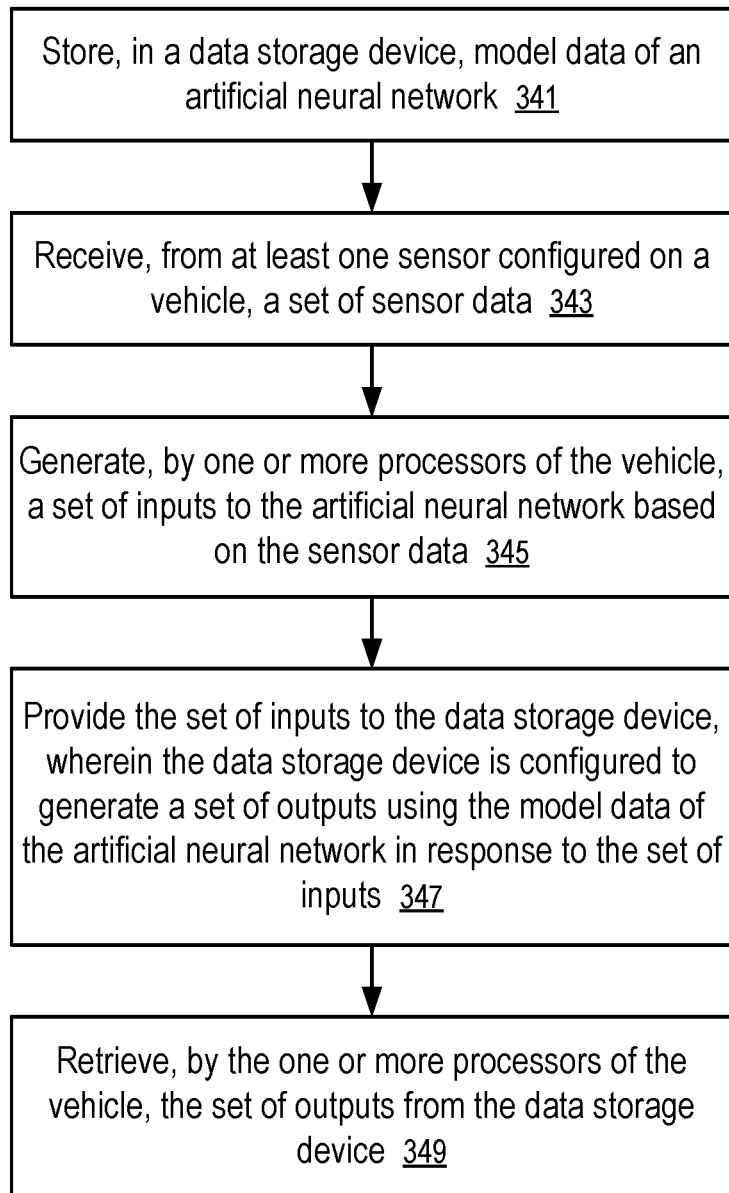


FIG. 19

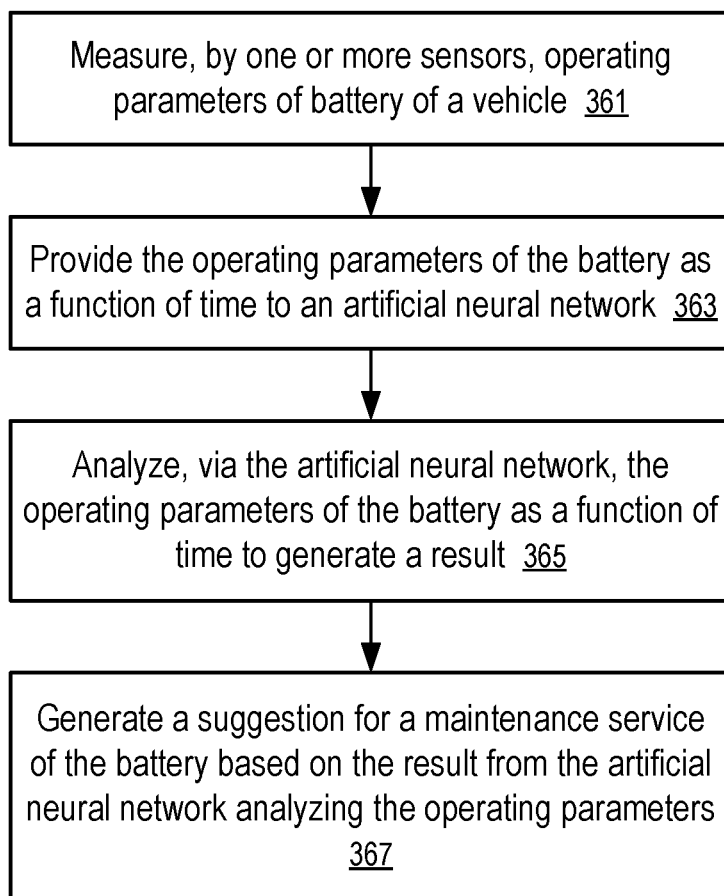
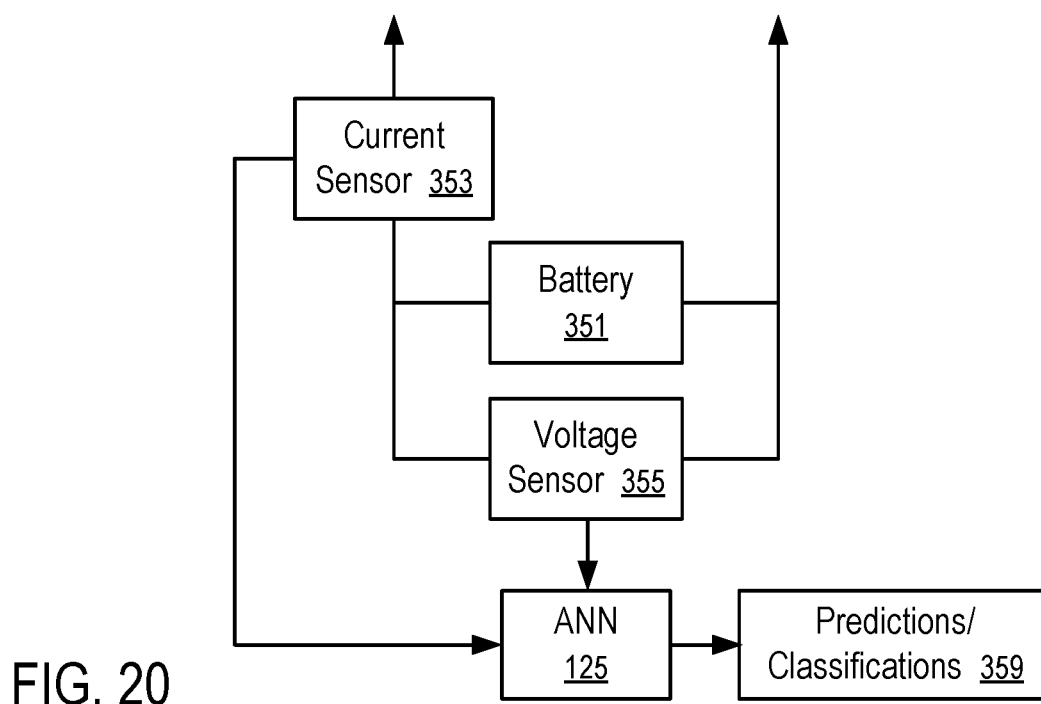


FIG. 21

PREDICTIVE MAINTENANCE OF AUTOMOTIVE BATTERY

FIELD OF THE TECHNOLOGY

[0001] At least some embodiments disclosed herein relate to maintenance services of vehicles in general and more particularly, but not limited to, predictive maintenance services of automotive battery.

BACKGROUND

[0002] Automotive maintenance is conventionally scheduled based on a predetermined milestone of operations. For example, a routine maintenance service can be scheduled once for every three or six months, or after a predetermined distance traveled (e.g., 3000 miles, 6000 miles, or 15000 miles).

[0003] When a component of a motor vehicle breaks down or malfunctions during the operation of the vehicle, such an incident can be a safety hazard. After such an incident occurs, a trip for the service of the vehicle would be scheduled soon at an inconvenient time.

[0004] Recent developments in the technological area of autonomous driving allow a computing system to operate, at least under some conditions, control elements of a motor vehicle without the assistance from a human operator of the vehicle.

[0005] For example, sensors (e.g., cameras and radars) can be installed on a motor vehicle to detect the conditions of the surroundings of the vehicle traveling on a roadway. A computing system installed on the vehicle analyzes the sensor inputs to identify the conditions and generate control signals or commands for the autonomous adjustments of the direction and/or speed of the vehicle, with or without any input from a human operator of the vehicle.

[0006] In some arrangements, when a computing system recognizes a situation where the computing system may not be able to continue operating the vehicle in a safe manner, the computing system alerts the human operator of the vehicle and requests the human operator to take over the control of the vehicle and drive manually, instead of allowing the computing system to drive the vehicle autonomously.

[0007] U.S. Pat. No. 9,533,579, entitled "Electronic Control Apparatus for Electrically-Driven Vehicle" and published Jan. 3, 2017, discloses an electronic control apparatus of a vehicle that has a self-diagnosis function.

[0008] Autonomous driving and/or advanced driver assistance system (ADAS) typically involves artificial neural network (ANN) for the identification of events and/or objects that are captured in sensor inputs.

[0009] In general, an artificial neural network (ANN) uses a network of neurons to process inputs to the network and to generate outputs from the network.

[0010] For example, each neuron in the network receives a set of inputs. Some of the inputs to a neuron may be the outputs of certain neurons in the network; and some of the inputs to a neuron may be the inputs provided to the neural network. The input/output relations among the neurons in the network represent the neuron connectivity in the network.

[0011] For example, each neuron can have a bias, an activation function, and a set of synaptic weights for its inputs respectively. The activation function may be in the

form of a step function, a linear function, a log-sigmoid function, etc. Different neurons in the network may have different activation functions.

[0012] For example, each neuron can generate a weighted sum of its inputs and its bias and then produce an output that is the function of the weighted sum, computed using the activation function of the neuron.

[0013] The relations between the input(s) and the output(s) of an ANN in general are defined by an ANN model that includes the data representing the connectivity of the neurons in the network, as well as the bias, activation function, and synaptic weights of each neuron. Using a given ANN model a computing device computes the output(s) of the network from a given set of inputs to the network.

[0014] For example, the inputs to an ANN network may be generated based on camera inputs; and the outputs from the ANN network may be the identification of an item, such as an event or an object.

[0015] A spiking neural network (SNN) is a type of ANN that closely mimics natural neural networks. An SNN neuron produces a spike as output when the activation level of the neuron is sufficiently high. The activation level of an SNN neuron mimics the membrane potential of a natural neuron. The outputs/spikes of the SNN neurons can change the activation levels of other neurons that receive the outputs. The current activation level of an SNN neuron as a function of time is typically modeled using a differential equation and considered the state of the SNN neuron. Incoming spikes from other neurons can push the activation level of the neuron higher to reach a threshold for spiking. Once the neuron spikes, its activation level is reset. Before the spiking, the activation level of the SNN neuron can decay over time, as controlled by the differential equation. The element of time in the behavior of SNN neurons makes an SNN suitable for processing spatiotemporal data. The connectivity of SNN is often sparse, which is advantageous in reducing computational workload.

[0016] In general, an ANN may be trained using a supervised method where the parameters in the ANN are adjusted to minimize or reduce the error between known outputs resulted from respective inputs and computed outputs generated from applying the inputs to the ANN. Examples of supervised learning/training methods include reinforcement learning, and learning with error correction.

[0017] Alternatively, or in combination, an ANN may be trained using an unsupervised method where the exact outputs resulted from a given set of inputs is not known before the completion of the training. The ANN can be trained to classify an item into a plurality of categories, or data points into clusters.

[0018] Multiple training algorithms can be employed for a sophisticated machine learning/training paradigm.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] The embodiments are illustrated by way of example and not limitation in the figures of the accompanying drawings in which like references indicate similar elements.

[0020] FIG. 1 shows a system in which a vehicle is configured with a data storage device to collect and process sensor data according to some embodiments.

[0021] FIG. 2 shows an autonomous vehicle having a data storage device according to one embodiment.

[0022] FIGS. 3-5 illustrate training of artificial neural networks for maintenance service prediction according to some embodiments.

[0023] FIG. 6 shows a method of predictive maintenance according to one embodiment.

[0024] FIG. 7 shows a data storage device to accelerate neural network computations according to one embodiment.

[0025] FIG. 8 shows a storage media component to accelerate neural network computations according to one embodiment.

[0026] FIG. 9 shows a method to accelerate neural network computations in motor vehicles according to one embodiment.

[0027] FIG. 10 shows a data storage device configured to support neural network computations according to one embodiment.

[0028] FIG. 11 illustrates the configuration of a namespace for an artificial neural network (ANN) model according to one embodiment.

[0029] FIG. 12 illustrates the configuration of a namespace for the inputs to artificial neurons according to one embodiment.

[0030] FIG. 13 illustrates the configuration of a namespace for the outputs from artificial neurons according to one embodiment.

[0031] FIGS. 14-16 show methods of predictive maintenance supported by a model partition, an input partition and an output partition according to one embodiment.

[0032] FIG. 17 shows communications with a data storage device to implement neural network computation according to one embodiment.

[0033] FIG. 18 shows communications within a data storage device to implement neural network computation according to one embodiment.

[0034] FIG. 19 shows a method of communicating with a data storage device to implement neural network computation according to one embodiment.

[0035] FIG. 20 shows the use of an artificial neural network (ANN) to monitor the health of automotive battery according to one embodiment.

[0036] FIG. 21 shows a method of predictive maintenance of battery of a vehicle according to one embodiment.

DETAILED DESCRIPTION

[0037] At least some embodiments disclosed herein provide systems, methods and apparatus to process sensor data generated in a motor vehicle, or another vehicle with or without an advanced driver assistance system (ADAS), to facilitate predictive maintenance.

[0038] Before a component of a motor vehicle breaks down or malfunctions during the operation of a vehicle, there can be indication of whether the component needs replacement or maintenance. Such indications may not be noticeable to a typical driver or passengers. However, sensor data can be collected and analyzed to predict the probability of component failures. The prediction can be used to schedule maintenance services, which can reduce or eliminate the chances of incidents where a component of a vehicle breaks down or malfunctions during the operation of the vehicle on a roadway. Further, the prediction allows the service trip to be scheduled at a convenient time.

[0039] For example, sensors can be installed in an automotive system to collect data during its routine operations; and the sensor data can be used to predict whether and how

soon a component needs replacement or maintenance. The sensor data can be provided as input to an artificial neural network (ANN) (e.g., spiking neural network (SNN)) of an artificial intelligent (AI) system to train itself (e.g., using an unsupervised machine learn technique) in a time period in which the vehicle is expected to operate normally. The training customizes the neural network for the specific operating environment(s) of the driver, passenger, or user of the vehicle and the personalized operating habits of the vehicle occupant(s). Subsequently, when the operating data deviates the normal mode, the artificial neural network can detect abnormal conditions. The AI system can be used to suggest a maintenance service and/or identify the component that likely needs replacement or maintenance.

[0040] FIG. 1 shows a system in which a vehicle is configured with a data storage device to collect and process sensor data according to some embodiments.

[0041] The system of FIG. 1 includes a vehicle (111) having a data storage device (101). Optionally, the vehicle (111) has an advanced driver assistance system (ADAS) (105) and one or more sensors (103) that provide sensor data input to the ADAS (105) and/or the data storage device (101). The data storage device (101) is configured to use an artificial neural network (ANN) (125) to predict/identify a need for a maintenance service based on the data collected by the sensors (103). The ADAS (105) can be omitted without impacting the predictive maintenance features. In some implementations, at least a portion of the data generated by the sensors (103) is used in both the ADAS (105) for driver assistance and in the ANN (125) for maintenance prediction. Optionally, the output of the ANN (124) can be used in both the data storage device (101) and in the ADAS (105).

[0042] The sensor(s) (103) can include digital cameras, lidars, radars, ultrasound sonars, brake sensors, speed sensors, acceleration sensors, airbag sensors, a GPS (global positioning system) receiver, audio sensors/microphones, vibration sensors, force/stress sensors, deformation sensors, motion sensors, temperature sensors, etc. Some of the sensors (103) can be configured primarily to monitor the environment of the vehicle (111); and other sensors (103) can be configured primarily to monitor the operating condition of one or more component of the vehicle (111), such as an internal combustion engine, an exhaust system, an electric motor, a brake, a tire, a battery, etc.

[0043] The outputs of the sensor(s) (103) as a function of time are provided as a sensor data stream to the ADAS (105) and/or the ANN (125) to provide driver assistance (e.g., autonomous driving) and maintenance prediction.

[0044] For example, the vehicle (111) can have a wireless communication device to communicate with a remote server (119) via wireless signals (113) and a communication network (117). The remote server (119) is typically configured at a location away from a road (102) on which the vehicle (111) is in service. For example, the vehicle (111) may provide some sensor data (121) to the server (119) and receive update of the ANN (125) from the server (119).

[0045] One example of the communication network (117) is a cell phone network having one or more base stations (e.g., 115) to receive the wireless signals (e.g., 113). Another example of the communication network (117) is internet, where the wireless local area network signals (e.g., 113) transmitted by the vehicle (113) is received in an access point (e.g., 115) for further communication to the server

(119). In some implementations, the vehicle (111) uses a communication link (107) to a satellite (109) or a communication balloon to communicate with the server (119).

[0046] The server (119) can also communicate with one or more maintenance service facilities (e.g., 127) to receive maintenance service data (123) of vehicles (e.g., 111). The maintenance service data (123) can include inspection records and/or service records of components of the vehicles (e.g., 111). For example, the inspection records and/or service records can indicate the degree of wear and tear of components inspected during their services at the maintenance service facilities (e.g., 127), the identification of failed or malfunctioning components, etc. The sensor data (121) of the vehicles (e.g., 111) in a time period prior to the services and the maintenance service data (123) can be used to train an ANN (125) to predict the probability of a component requiring a maintenance service. The updated ANN (125) can be used to predict and suggest a maintenance service for a vehicle (111) based on sensor data (121) received in a recent period of time. Alternatively, the update ANN (125) can be transmitted to the vehicle (111); and the vehicle (111) can use the data generated from the sensors (103) during routine operations of the vehicle (111) to predict and suggest a maintenance service.

[0047] The data storage device (101) of the vehicle (111) can be configured to record sensor data for a period of time that can be used in the ANN for predictive maintenance. Maintenance prediction is typically for a relative long period of time (e.g., a few days, weeks and/or months). In contrast, sensor data recorded for the review of an accident, collision, or near collision involving an autonomous vehicle is typically for a short period of time (e.g., 30 seconds to a few minutes). Thus, a typical black box data recorder configured to record sensor data for the review/analysis of an accident or collision is insufficient for predictive maintenance.

[0048] Optionally, the data storage device (101) stores the sensor data of a period of time leading to a trip to a maintenance service facility (e.g., 127). The maintenance service facility (e.g., 127) can download the sensor data (121) from the data storage device (101) and provide the sensor data (121) and the corresponding maintenance service data (123) to the server (119) to facilitate the training of the ANN (125).

[0049] Optionally, or in combination, the data storage device (101) is configured with a machine learning module to customize and/or train the ANN (125) installed in the vehicle (111) for predictive maintenance.

[0050] For example, the machine learning module of the data storage device (101) can be used to calibrate the ANN (125) to account for the typical/daily environment in which the vehicle (111) is being operated and/or driving preferences/habits of the driver(s) of the vehicle (111).

[0051] For example, during a period of time when the vehicle is expected to be operated under typical/daily environment with healthy components, the sensor data generated by the sensors (103) can be used to train the ANN (125) to recognize the patterns of sensor data that represents trouble free operations. Such patterns can vary for different vehicles (e.g., 111) based on their routine operating environments and the driving habits/characteristics of their drivers. The training allows the ANN (125) to detect deviations from the recognized normal patterns and report anomaly for maintenance predictions.

[0052] For example, the ANN (125) can include an SNN configured to classify time-based variations of sensor data and/or detect deviation from known patterns of sensor data of the vehicle (111) operated in the normal/healthy condition but in a personalized environment (e.g., a daily route of a driver/passenger) and/or operated under a personalized driving habit/pattern.

[0053] FIG. 2 shows an autonomous vehicle (111) having a data storage device (101) according to one embodiment. For example, the vehicle (111) in the system of FIG. 1 can be implemented using the autonomous vehicle (111) of FIG. 2.

[0054] The vehicle (111) of FIG. 2 is configured to have an advanced driver assistance system (ADAS) (105). The ADAS (105) of the vehicle (111) can have an Artificial Neural Network (ANN) (125) for object detection, recognition, identification, and/or classification. The ANN (125) and/or another neural network (e.g., configured in the data storage device (101)) can be used to predict the probability of a component of the vehicle (111) requiring a maintenance service (e.g., repair, replacement, or adjustment).

[0055] Preferably, the data storage device (101) is configured to process sensor data at least partially for predictive maintenance with reduced computation burden on the processors (133) that are tasked to operate the ADAS (105) and/or other components, such as an infotainment system (149).

[0056] The vehicle (111) typically includes an infotainment system (149), a communication device (139), one or more sensors (103), and a computer system (131) that is connected to some controls of the vehicle (111), such as a steering control (141) for the direction of the vehicle (111), a braking control (143) for stopping of the vehicle (111), an acceleration control (145) for the speed of the vehicle (111), etc. In some embodiments, the vehicle (111) in the system of FIG. 1 has a similar configuration and/or similar components.

[0057] Some of the sensors (103) are required for the operations of the ADAS (105); and some of the sensors (103) are used to collect data related to the health of the components of the vehicle (111), which may not be used in the ADAS (105). Optionally, the sensor data generated by the sensors (103) can also be used to predict the likelihood of imminent failure of a component. Such a prediction can be used in the ADAS (105) to take emergency actions to render the vehicle in a safe state (e.g., by reducing speed and/or pulling off to park).

[0058] The computer system (131) of the vehicle (111) includes one or more processors (133), a data storage device (101), and memory (135) storing firmware (or software) (147), including the computer instructions and data models for ADAS (105).

[0059] The one or more sensors (103) of the vehicle can include a visible light camera, an infrared camera, a lidar, radar, or sonar system, a peripheral sensor, a global positioning system (GPS) receiver, a satellite positioning system receiver, a brake sensor, and/or an airbag sensor. Further, the sensors (103) can include audio sensors (e.g., microphone) configured to monitor noises from various components and locations in the vehicle (111), a vibration sensor, a pressure sensor, a force sensor, a stress sensor, and/or a deformation sensor configured to measure loads on a component of the vehicle (111), accelerometers and/or gyroscope sensors measuring the motions of some components of the vehicle (111),

etc. Such sensors can be used to monitor the operating status and/or health of the components for predictive maintenance.

[0060] The sensor(s) (103) can provide a stream of real time sensor data to the computer system (131). The sensor data generated by a sensor (103) of the vehicle (111) can include an image that captures an object using a camera that images using lights visible to human eyes, or a camera that images using infrared lights, or a sonar, radar, or LIDAR system. Image data obtained from at least one sensor of the vehicle is part of the collected sensor data for recording in the data storage device (101) and/or as input to the ANN (125). For example, a camera can be used to obtain roadway information for the travel of the vehicle (111), which can be processed by the ANN (125) to generate control signals for the vehicle (111). For example, a camera can be used to monitor the operation state/health of a component of the vehicle (111), which can be processed by the ANN (125) to predict or schedule a maintenance service.

[0061] The sensor data generated by a sensor (103) of the vehicle (111) can include an audio stream that captures the characteristics of sounds at a location on the vehicle (111), such as a location near an engine, a motor, a transmission system, a wheel, a door, a window, etc. The audio data obtained from at least one sensor (103) of the vehicle (111) can be part of the collected sensor data for recording in the data storage device (101) and/or as input to the ANN (125). For example, the audio stream can be used to monitor the operation state/health of a component of the vehicle (111) (e.g., an internal combustion engine, an exhaust system, an electric motor, a brake), which can be processed by the ANN (125) to predict or schedule a maintenance service.

[0062] The infotainment system (149) can be used to present the predicted or schedule maintenance service. Optionally, the communication device (139) can establish a connection to a mobile device of the driver of the vehicle (111) to inform the driver of the recommended maintenance service and/or the recommended data of the service, to calendar the appointment, etc.

[0063] When the vehicle (111) is configured with an ADAS (105), the outputs of the ADAS (105) can be used to control (e.g., (141), (143), (145)) the acceleration of the vehicle (111), the speed of the vehicle (111), and/or the direction of the vehicle (111), during autonomous driving.

[0064] FIGS. 3-5 illustrate training of artificial neural networks for maintenance service prediction according to some embodiments.

[0065] In FIG. 3, a module (171) of supervised machine learning is used to train an artificial neural network (125) to minimize the differences between the service prediction (129) generated from the sensor data (121) and the maintenance service data (123).

[0066] For example, the maintenance service data (123) can identify the measured wear and tear of a component as a function of time to predict a time to a recommended service. The sensor data (121) can be used in the ANN (125) to generate a predicted time to the recommended service. The supervised machine learning module (171) can adjust the artificial neural network (125) to reduce/minimize the difference between the time predicted based on the sensor data (121) and the time computed from the measurement of wear and tear.

[0067] For example, the maintenance service data (123) can identify a component that is replaced or repaired in the maintenance service facility (127). The sensor data (121)

recorded for a time period prior to the replacement or repair of the component can be used to calculate the times to the replacement or repair. Further, the segments of the stream of sensor data in the time period before the replacement or repair can be used in the ANN (125) to generate a prediction to the time of the replacement or repair. The supervised learning (171) can be used to adjust the ANN (125) to reduce the predicted time to the replacement or repair and the actual time to the replacement or repair.

[0068] The supervised learning (171) of FIG. 2 can be applied in the server (119) based on the sensor data of a population of vehicles and their maintenance service data (123) to generate a generic ANN for the population of the vehicles.

[0069] The supervised learning (171) of FIG. 2 can be applied in the vehicle (111) based on the sensor data of the vehicle and its maintenance service data (123) to generate a customized/personalized ANN for the population of the vehicles. For example, a generic ANN can be initially used in the vehicle (111); and the sensor data of the vehicle (111) and its maintenance service data (123) can be used to further train the ANN (125) of the vehicle for customization/personalization of the ANN (125) in the vehicle (111).

[0070] In FIG. 4, a module (175) of unsupervised machine learning is used to train or refine an artificial neural network (125) to facilitate anomaly detection (173). The unsupervised machine learning module (175) is configured to adjust the ANN (e.g., SNN) the classification, clustering, or recognized pattern in the sensor data (121) such that a degree of deviation from the classification, clustering, or recognized pattern in the sensor data (121) generated in a recent time period can be used to signal the detection (173) of anomaly. Anomaly detection (173) allows the vehicle (111) to be scheduled for inspection in a maintenance service facility (127). Optionally, after the inspection, maintenance service data (123) can be used to apply a supervised learning (171) to generate more precise predictions to a service, as in FIG. 3.

[0071] Typically, a vehicle (111) can be assumed to be operating in a normal/healthy condition in a certain time period. For example, after a new vehicle (111) is initially delivered for service, the vehicle (111) can be assumed to provide trouble-free services for at least a period of time (e.g., a few months). For example, after a time period following the replacement or repair of a component, the component can be assumed to provide trouble-free service for at least a period of time (e.g., a few months or a year). Thus, the sensor data (121) obtained during this period of time can be pre-classified as “normal” to train the ANN (125) using an unsupervised learning (175) as in FIG. 4, or a supervised learning (171) as in FIG. 5.

[0072] For example, the sensor data (121) collected via during the “normal” service time period of the vehicle (111) or a component can be classified via an unsupervised learning (175) into a number of clusters. Different clusters may correspond to different types of normal conditions (e.g., traveling on different routes, on roads with different surface conditions, on days with different weather conditions, in different time periods of a day, different days in a week, different mood of driving habits of the driver). When a subsequent sensor data (121) is classified outside of the “normal” clusters, an anomaly is detected.

[0073] Optionally, a supervised machine learning (171) can be used to train the ANN (125), as illustrated in FIG. 5.

During the “normal” service period of the vehicle (111) or a component, an expected classification (177) can be used to label the sensor data (121). The supervised learning (171) can be used to minimize the classification differences between the predictions (179) made using the ANN (125) according to the sensor data (121) and the expected classification (177). Further, when the sensor data (121) is known to be “abnormal” (e.g., after a diagnosis made in the maintenance service facility (127) or by the user, driver, or passenger of the vehicle (111)), the expected classification (177) can be changed to “abnormal” for further training of the ANN (125) for direct recognition of the anomaly (e.g., instead of relying upon deviation from known “normal” clusters for an inference of anomaly).

[0074] Thus, the ANN (125) can be trained to identify abnormal sensor data and estimate the degree of severity in anomaly to schedule a maintenance service.

[0075] FIG. 6 shows a method of predictive maintenance according to one embodiment. For example, the method of FIG. 6 can be implemented in the data storage device (101) of FIG. 1 or 2 in the vehicle (111) or in a computer system (131) in the vehicle (111) of FIG. 2.

[0076] At block 201, a sensor (e.g., 103) installed in a vehicle (111) generates a sensor data stream (e.g., 121) during operations of the vehicle (111) on a road (102).

[0077] At block 203, the sensor data stream (e.g., 121) is provided into an artificial neural network (ANN) (125). For example, the ANN (125) can include a spiking neural network (SNN).

[0078] At block 205, the artificial neural network (ANN) (125) generates, based on the sensor data stream (e.g., 121), a prediction of a maintenance service.

[0079] At block 207, a data storage device (101) configured on the vehicle stores at least a portion of the sensor data stream (e.g., 121).

[0080] At block 209, the artificial neural network (ANN) is trained using the sensor data stream (e.g., 121) collected within a predetermined time period from the vehicle leaving a factory or a maintenance service facility (127).

[0081] For example, the artificial neural network (ANN) can be configured to identify a component of the vehicle (111) that needs repair or replacement in the maintenance service and/or identify a predicted time period to failure or malfunctioning of the component, or a suggested time period to a recommended maintenance service of the component prior to the failure or malfunctioning of the component. Thus, the performance of the predicted maintenance service can avoid an incident of failure or malfunctioning of the component while the vehicle (111) operates on the road (102).

[0082] For example, the sensor (103) can be a microphone mounted in vicinity of the component, a vibration sensor attached to the component, a pressure sensor installed in the component, a force or stress sensor mounted on or attached to the component, a deformation sensor attached to the component, an accelerometer configured to measure motion parameters of the component.

[0083] Optionally, the data storage device (101), the computer system (131) of the vehicle (111), and/or a server (119) remote from the vehicle can have a machine learning module configured to train the artificial neural network (ANN) (125) during a period of time in which the vehicle (111) is assumed to be in a healthy state, such as a prede-

termined time period from the vehicle (111) leaving a factory or a maintenance service facility (127).

[0084] For example, the machine learning module can use an unsupervised machine learning (175) to train the ANN (125) to recognize/classify normal patterns of sensor data (121) and thus to have the capability to detect anomaly based on deviation from the normal patterns, as illustrated in FIG. 4. Alternatively, supervised machine learning (171) can be used, as illustrated in FIG. 3 or 5.

[0085] For example, unsupervised machine learning (175) can be applied by the data storage device (101) or the computer system (131) of the vehicle (111) during the predetermined period of time in which the vehicle and/or the component is known to be operating without troubles or degradations.

[0086] Alternatively, or in combination, some of the sensor data (121) stored in the data storage device (101) of the vehicle (111) can be uploaded to the server (119) for training the ANN (125).

[0087] In at least some embodiments disclosed herein, the data storage device (101) is configured to accelerate the computations of the artificial neural network (ANN) (125) of the vehicle (111).

[0088] For example, in addition to the typical operations to support data access and storage, the data storage device (101) can be further configured to perform at least part of the computations involving the artificial neural network (ANN) (125), such as the generation of the predictions (e.g., 129 or 173) or classifications (e.g., 179) from the sensor data (121) and/or the adjustment of the ANN (125) through unsupervised machine learning (175) (e.g., as illustrated in FIG. 4) and/or supervised machine learning (171) (e.g., as illustrated in FIG. 3 or 5).

[0089] For example, the computations configured in the data storage device (101) can be used to reduce the amount of data to be transmitted to the processor(s) (133) to use or apply the ANN (125) and/or reduce the computation tasks of the processor(s) (133) in evaluating the outputs of the ANN (125) and/or in training the ANN (125). Such an arrangement can result in faster output from the data storage device (101) and/or lower energy usage, since the data would not have to be moved in and out of the memory to a dedicated, standalone neural network accelerator. The computation capability of the data storage device (101) in processing data related to the ANN (125) enables the computer system (131) of the motor vehicle (111) to monitor the health of the automotive components (e.g., in a non-real-time manner, or a pseudo-real-time manner), with reduced impact, or no impact, on the processing of mission critical tasks (e.g., autonomous driving by the ADAS (105)). Further, the computation capability of the data storage device (101) can be used to accelerate the processing of the sensor data for ADAS (105) and thus improve the processing of mission critical tasks.

[0090] FIG. 7 shows a data storage device (101) to accelerate neural network computations according to one embodiment. For example, the data storage device (101) of FIG. 7 can be used to implement the data storage device (101) of the vehicle (111) illustrated in FIG. 1 or 2.

[0091] In FIG. 7, the data storage device (101) has a host interface (157) configured to communicate with a processor (e.g., 133). For example, the communication between the processor (e.g., 133) and the host interface (157) can be, at least in part, in accordance with a communication protocol

for a peripheral component interconnect express (PCIe) bus, a serial advanced technology attachment (SATA) bus, a universal serial bus (USB) bus, and/or a storage area network (SAN).

[0092] For example, the host interface (157) can be used to receive sensor data (121) generated by the sensors (103) of the vehicle (111) to optionally store a portion of the sensor data (121) in the storage media components (161 to 163).

[0093] For example, each of the storage media components (161 to 163) can be a memory integrated circuit configured to store data. For example, a media component (161 or 163) can include one or more integrated circuit dies embedded in an integrated circuit package. An integrated circuit die can have many memory units formed thereon to store data.

[0094] In general, some memory integrated circuits are volatile and require power to maintain the stored data; and some memory integrated circuits are non-volatile and can retain the stored data even when not powered.

[0095] Examples of non-volatile memory include flash memory, memory units formed based on negative-and (NAND) logic gates, negative-or (NOR) logic gates, phase-change memory (PCM), magnetic memory (MRAM), resistive random-access memory, cross point storage and memory devices. A cross point memory device uses transistor-less memory elements, each of which has a memory cell and a selector that are stacked together as a column. Memory element columns are connected via two perpendicular lays of wires, where one lay is above the memory element columns and the other lay below the memory element columns. Each memory element can be individually selected at a cross point of one wire on each of the two layers. Cross point memory devices are fast and non-volatile and can be used as a unified memory pool for processing and storage. Further examples of non-volatile memory include Read-Only Memory (ROM), Programmable Read-Only Memory (PROM), Erasable Programmable Read-Only Memory (EPROM) and Electronically Erasable Programmable Read-Only Memory (EEPROM) memory, etc. Examples of volatile memory include Dynamic Random-Access Memory (DRAM) and Static Random-Access Memory (SRAM).

[0096] The data storage device (101) can have a controller (151) that includes volatile local memory (153) and at least one processing device (155).

[0097] The local memory of the controller (151) can be an embedded memory configured to store instructions for performing various processes, operations, logic flows, and routines that control operation of the processing device (155), including handling communications between the data storage device (101) and the processor(s) (e.g., 133) of the vehicle (111), and other functions described herein. The local memory (151) of the controller (151) can include read-only memory (ROM) for storing micro-code and/or memory registers storing, e.g., memory pointers, fetched data, etc., and/or volatile memory, such as Dynamic Random-Access Memory (DRAM) and Static Random-Access Memory (SRAM).

[0098] In FIG. 7, the data storage device (101) includes a neural network accelerator (159) coupled to the controller (151) and/or the storage media components (161 to 163).

[0099] For example, the neural network accelerator (159) can be configured to perform matrix arithmetic computations. The computations involving ANN (125) have matrix

multiplication and accumulation operations, which can be computational intensive for a generic processor (e.g., 133). Using the neural network accelerator (159) to perform the matrix arithmetic computations can reduce the data to be transmitted to the processor(s) (133) of the vehicle (111) and reduce the computation workload for the processor(s) (133).

[0100] For example, when the ANN (125) includes a spiking neural network (SNN), the simulation of the differential equation(s) for controlling the activation level of SNN neurons can be computationally intensive for a generic processor (e.g., 133). The neural network accelerator (159) can use special hardware to simulate the differential equation(s) and thus improve the computational efficiency of the computer system (131) as a whole.

[0101] In some implementations, the neural network accelerator (159) is an integrated circuit device separate from the controller (151) and/or the storage media components (161 to 163). Alternatively, or in combination, a neural network accelerator (159) is integrated with the controller (151) in an integrated circuit package. Further, a neural network accelerator (159) can be integrated in at least one of the storage media components (161 to 163), as illustrated in FIG. 8.

[0102] FIG. 8 shows a storage media component (160) to accelerate neural network computations according to one embodiment. For example, each or some of the storage media components (161 to 163) in FIG. 7 can be implemented using a storage media component (160) of FIG. 8.

[0103] In FIG. 8, the storage media component (160) can be housed within an integrated circuit package. An input/output (I/O) interface (171) of the storage media component (160) is configured to process input/output signals in the pins of the integrated circuit package. For example, the input/output signals can include address signals to specify locations in the media units (175) and data signals representing data to be written in the media units (175) at the locations specified via the address signals, or data retrieved from the locations in the media units (175).

[0104] In FIG. 8, a neural network accelerator (159) is coupled with the control logic (173) and/or the media units (175) to perform computations that are used in the evaluation of the output of an ANN (125) and/or in the training of the ANN (125).

[0105] For example, the input/output interface (171) can receive addresses that identify matrices that are stored in the media units and that are to be operated upon via the neural network accelerator (159). The storage media component (160) can provide the computation results of the neural network accelerator (159) as the output data responsive to the addresses, store the output data in a buffer for further operations, store the output data into a location in the media units (175) specified via the address signals. Thus, the computations performed by the neural network accelerator (159) can be within the storage media component (160), which is close to the media units (175) in which the matrix data is stored. For example, each of the media units (175) can be an integrated circuit die on which memory units of non-volatile memory are formed.

[0106] For example, the state data of SNN neurons can be stored in the media units (175) according to a predetermined pattern. The neural network accelerator (159) can automatically update the states of the SNN neurons according to the differential equation(s) for controlling the activation level of SNN neurons over time. Optionally, the neural network

accelerator (159) is configured to processing spiking of neurons in the neural network. Alternatively, the neural network accelerator (159) of the data storage device (101) and/or the processor(s) (133) can be configured to process the spiking of neurons and/or accumulation of inputs to the SNN.

[0107] FIG. 9 shows a method to accelerate neural network computations in motor vehicles according to one embodiment. For example, the method of FIG. 9 can be implemented in a vehicle (111) of FIG. 1 or 2 using a data storage device (101) of FIG. 7 and/or a storage media component (160) of FIG. 8. For example, the method of FIG. 9 can be used in combination with the method of FIG. 6.

[0108] At block 221, a data storage device (101) of a vehicle (111) receives a sensor data stream from at least one sensor (e.g., 103) configured on the vehicle (111).

[0109] At block 223, the data storage device (101) stores at least a portion of the sensor data stream.

[0110] At block 225, a neural network accelerator (159) configured within the data storage device (101) performs at least a portion of computations based on an artificial neural network (125) and the sensor data stream.

[0111] At block 227, a maintenance service of the vehicle (111) is predicted based at least in part on the computations performed by the neural network accelerator (159) configured within the data storage device (101).

[0112] Optionally, at block 229, the artificial neural network (ANN) is trained in the vehicle (111), using at least in part the neural network accelerator and using the sensor data stream collected within a predetermined time period, such as a period following the delivery of the new vehicle (111) from a factory or following the replacement of a component in a maintenance service facility (127).

[0113] For example, the neural network accelerator (159) can be configured on an integrated circuit device that is separate from a controller (151) of the data storage device and/or separate from the storage media components (161 to 163).

[0114] For example, the neural network accelerator (159) can be configured on an integrated circuit device that includes a controller (151) of the data storage device (101), or on an integrated circuit device that includes storage media component (160, 1061 or 163) of the data storage device (101).

[0115] For example, the neural network accelerator (159) can be configured to perform computations, such as matrix arithmetic computations for ANN and/or or differential equation simulations for SNN, using data stored in the data storage device (101).

[0116] Examples of the matrix arithmetic computations include matrix multiplication and accumulation operations. After a computation to generate a result of the matrix arithmetic computations using a data stored in the data storage device (101), the neural network accelerator (159) can provide the result as output of the data storage device (111) in retrieving data (e.g., in response to a read command). Alternatively, or in combination, the result of the matrix arithmetic computation can be buffered in the data storage device (101) as operand for a next matrix computation performed in combination with a matrix of data retrieved from the non-volatile memory via a read command received in the host interface (157).

[0117] When the artificial neural network (ANN) (125) includes a spiking neural network (SNN), the neural net-

work accelerator can be configured to simulate a differential equation controlling activation levels of neurons in the spiking neural network (SNN). Optionally, the storage media component is configured to store states of the neurons in the spiking neural network according to a predetermined pattern; and the neural network accelerator is configured to automatically update the states of the neurons over time according to the differential equation. For example, the neural network accelerator (159) can be configured to train the spiking neural network (SNN) via unsupervised machine learning to detect anomaly.

[0118] The computations performed by the neural network accelerator (159) according to an artificial neural network (ANN) (125) involve different types of data that have different patterns of usages of the data storage device (101).

[0119] For example, making a prediction using the artificial neural network (ANN) (125) includes the use of data specifying the model of the artificial neural network (ANN) (125), input data provided to the artificial neurons, and output data generated by the artificial neurons.

[0120] The storage capacity of the data storage device (101) can be partitioned into different portions for the different types of ANN-related data. The different portions can be separately configured to optimize the access and storage of the corresponding data according to their patterns of usages by the neural network accelerator (159) and/or the processor(s) (133) of the computer system (131) in which the data storage device (101) is configured.

[0121] The model of the artificial neural network (ANN) (125) can include the parameters specifying the static attributes of individual artificial neurons in the ANN (125) and the neuron connectivity in the ANN (125). The model data of the ANN (125) is static and does not change during the prediction calculation made using the ANN (125). Thus, the usage pattern of the model data is mostly read. However, the model data of the ANN (125) can change when an updated ANN (125) is installed. For example, the vehicle (111) can download an updated ANN (125) from the server (119) to the data storage device (101) of the vehicle (111) to update its prediction capability. The model data of the ANN (125) can also change during or after the training of the ANN (125) using a machine learning technique (e.g., 171 or 175). It is preferred to configure a separate partition or namespace of the data storage device (101) to store the model data, where the partition or namespace is operated according to configuration parameters that optimize the memory units for the specific usage patterns of the model data (e.g., mostly read, infrequent update). For example, when the memory units are implemented using a flash memory based on NAND logic gates, the memory units in the ANN model partition/namespace can be configured to operate in a multi-level cell (MLC) mode, a triple level cell (TLC) mode, or a quad-level cell (QLC) mode, wherein each memory cells stores two, three, or four bits for increased storage capability.

[0122] Input data provided to the artificial neurons in the ANN (125) can include external inputs and internal inputs. The external inputs are generated typically by the sensors (103) of the vehicle (111) but not by artificial neurons in the ANN (125). The external inputs can be saved in a cyclic fashion so that the input data of the most recent time period of a predetermined length of driving can be found in the data storage device (101). Thus, it is preferred to configure a separate partition or namespace of the data storage device (101) to store the external input data, where the partition or

namespace is operated according to configuration parameters that optimize the memory units for the storage pattern of the external input data (e.g., enhanced endurance, cyclic overwrite). For example, when the memory units are implemented using a flash memory based on NAND logic gates, the memory units in the ANN input partition/namespace can be configured to operate in a single level cell (SLC) mode, where each memory cell stores one bit of data for improved endurance in cyclic overwriting operations.

[0123] In some implementations, artificial neurons can have state variables that change over time in response to inputs during prediction calculations. For example, the activation level of a spiking neuron can change over time and is considered a dynamic state variable of the spiking neuron. In some implementations, such state variable data of artificial neurons has a similar storage usage pattern as the external input data; and thus, the state variable data can be stored in the partition or namespace configured for the external input data. In other implementations, the state variable data of artificial neurons is kept in a buffer and stored less frequently than the external inputs; and thus, another partition/namespace can be configured for storing the dynamic state variable data of artificial neurons.

[0124] Output data generated by the artificial neurons in the ANN (125) can be buffered for further access by the neural network accelerator (159) and/or the processor(s) (133) of the computer system (131). The output data can include external outputs and internal outputs. The external inputs are generated by artificial neurons as the output from the ANN (125), such as the results of classifications or predictions made by the ANN (125). The output of the ANN (125) is typically further processed by the processor(s) (133) of the computer system (131). The external inputs may be saved periodically (e.g., in a way similar to the storing of the state variable data). The internal outputs and/or some of the external outputs can be internal inputs to artificial neurons in the ANN (125). In general, it may not be necessary to store the internal outputs from the buffer of the data storage device to the storage media components. In some implementations, when the buffer capability of the data storage device (101) is insufficient to hold the entire state variable data and/or the internal outputs, the data storage device (101) can use a swap partition/namespace to extend the capacity of the buffer. The swap partition/namespace can be configured for optimized random access and for improved endurance.

[0125] External outputs and/or dynamic states of neurons can be saved in a separate output partition or namespace, in a cyclic way so that the external output data and/or dynamic states of the neurons can be periodically stored, and the most recent sets of the external outputs and/or dynamic states can be found in the data storage device (101). External outputs and/or dynamic states of neurons can be stored selectively, since some of such data can be re-generated by the ANN from the external inputs stored in the input partition or namespace. Preferably, the output partition or namespace is configured to store one or more sets of external outputs and/or dynamic states that cannot be created from the external inputs stored in the input partition or namespace. In storing data in a cyclic way in an input/output partition or namespace, the oldest stored data sets are erased to make room for the most recent data sets. The ANN input/output partition/namespace can be configured for an optimized sequential write stream for copying data from the buffer of

the data storage device into the memory units in the storage media components of the data storage device.

[0126] FIG. 10 shows a data storage device (101) configured to support neural network computations according to one embodiment. For example, the data storage device (101) can be used in the vehicle (111) in FIG. 1 or 2 to facilitate predictive maintenance and/or support the ADAS (105).

[0127] The data storage device (101) of FIG. 10 includes a host interface (157) and a controller (151), similar to the data storage device (101) of FIG. 7.

[0128] The storage capacity (181) of the data storage device (101) of FIG. 10 can be implemented using a set of storage media components, similar to the storage media components (161 to 163) in the data storage device (101) of FIG. 7.

[0129] A set of namespaces (183, 185, 187, . . .) can be created on the storage capacity (181) of the data storage device (101). Each of the namespace (e.g., 183, 185, or 187) corresponds to a named portion of the storage capacity (181). Logical addresses are defined within each namespace. An address map (191) is configured to map between the logical addresses defined in the namespaces (183, 185, 187, . . .) to the physical addresses of memory units in the storage media components (e.g., 161 to 163 illustrated in FIG. 7).

[0130] The address map (191) can include namespace optimization settings (192) for the namespaces (183, 185, and 187).

[0131] For example, an ANN model namespace (183) can be a memory/storage partition configured for the model data of the artificial neural network (ANN) (125). The namespace optimization settings (192) optimizes the memory operations in the ANN model namespace (183) according to the data usage pattern of ANN models (e.g., mostly read, infrequent update centric).

[0132] For example, a neuron input namespace (185) can be a memory/storage partition configured for the external input data to the artificial neural network (ANN) (125). The namespace optimization settings (192) optimizes the memory operations in the neuron input namespace (185) according to the data usage pattern of the external input data (e.g., for enhanced endurance supporting cyclic overwrite of continuous input data flow for sequential writes).

[0133] For example, a neuron output namespace (187) can be a memory/storage partition/configured for the external output data provided from the artificial neural network (ANN) (125). The namespace optimization settings (192) optimizes the memory operations in the neuron output namespace (187) according to the data usage pattern of the external output data (e.g., improved endurance for periodically overwrite of data with random read/write access).

[0134] The data storage device (101) includes a buffer (152) configured to store temporary/intermediate data of the artificial neural network (ANN) (125), such as the internal inputs/outputs of the artificial neurons in the ANN (125).

[0135] Optionally, a swap namespace can be configured in the storage capacity (181) to extend the capacity of the buffer (152).

[0136] Optionally, the address map (191) includes a mapping between logic memory addresses received in the host interface (157) to access data of artificial neurons and the identities of the artificial neurons. Thus, a read or write command to access one type of data of an artificial neuron

in one namespace can cause the controller 151 to access another type of data of the artificial neuron in another namespace.

[0137] For example, in response to a request to write external input data for a neuron into the storage capacity (181) of the data storage device (185), the address map (191) can be used to calculate the addresses of the model parameters of the neuron in the ANN model namespace (183) and read the model parameters into the buffer (152) to allow the neural network accelerator (159) to perform the computation of the output of the neuron. The output of the neuron can be saved in the buffer (152) as the internal input to other neurons (e.g., to reduce write amplification). Further, the identities of the other neurons connected to the neuron can also be retrieved from the ANN model namespace (183) into the buffer (152), which allows the neural network accelerator (159) and/or the processor to further process the propagation of the output in the ANN (125). The retrieval of the model data from the ANN model namespace (183) can be performed in parallel with the storing of the external input data into the neuron input namespace (185). Thus, the processors (133) of the computer system (131) of the vehicle (111) do not have to explicitly send in read commands for the retrieval of the model data from the ANN model namespace (183).

[0138] Similarly, in response to reading output data of a neuron, the address map (191) can be used to compute the addresses of the model parameters of the neuron stored in the ANN model namespace (183) and read the model parameters into the buffer (152) to allow the neural network accelerator (159) to apply internal inputs in the buffer (152) to the perform the computation of the output of the neuron. The computed output can be provided as a response to the reading of the output data for the neuron, without the data storage device (101) having to store the output data in the storage media components (e.g., 161 to 163). Thus, the processors (133) and/or the neural network accelerator (159) can control the computations of the neuron via writing inputs to neurons and/or reading outputs from neurons.

[0139] In general, incoming external input data to the ANN (125) can be raw sensor data (121) generated directly by the sensors (103) without processing by the processors (133) and/or the neural network accelerator (159). Alternatively, indirect sensor data (121) that has processed by the processors (133) for the ANN (125) from the signals from the sensors (103) can be provided as the external input data. The incoming external input data can be accepted in the host interface (157) and written in a cyclic way into the neuron input namespace (185), and automatically buffered in the buffer (152) for neural network accelerator (159) to generate neuron outputs using the model stored in the ANN model namespace (183). The outputs generated by the neural network accelerator (159) can be further buffered as internal inputs for further application of the model in the ANN model namespace (183). When the external outputs become available, the data storage device (101) can report the completion of the write requests with an indication of the availability of the external outputs. Optionally, the controller 151 and/or the neural network accelerator (159) can generate internal read commands to propagate signals in the ANN (125) in generating the external outputs. Alternatively, the host processors (133) can control the propagation of signals in the ANN (125) by selectively reading outputs of neurons; and

the data storage device (101) can actively buffer data that may be needed in the buffer (152) to accelerate the ANN computation.

[0140] FIG. 11 illustrates the configuration of a namespace (183) for an artificial neural network (ANN) model according to one embodiment. For example, the configuration of FIG. 11 can be implemented in the data storage device (101) illustrated in FIGS. 7 and/or 10. For example, the settings (193) of FIG. 11 can be part of the namespace optimization settings (192) of FIG. 10.

[0141] The configuration of FIG. 11 maps an ANN model namespace (183) to at least one storage media component A (161). Preferably, the at least one storage media component A (161) can be used by the controller (151) in parallel with storage media components (e.g., 163) that hosts the other namespaces (e.g., 185 and 187) of ANN data. For example, the storage media component A (161) can be in an integrated circuit package that is separate from the integrated circuit packages for the other namespaces (e.g., 185 and 187). Alternatively, the storage media components (161 to 163) are formed on separate integrated circuit dies embedded in a same integrated circuit package. Alternatively, the storage media components (161 to 163) can be formed on separate regions of an integrated circuit die, where the separate regions can be operated substantially in parallel (e.g., for read, for erase, and/or for write).

[0142] In FIG. 11, the settings (197) are optimized to the usage pattern of mostly read and infrequent update.

[0143] FIG. 12 illustrates the configuration of a namespace (185) for the inputs to artificial neurons according to one embodiment. For example, the configuration of FIG. 11 can be implemented in the data storage device (101) illustrated in FIGS. 7 and/or 10. For example, the settings (195) of FIG. 11 can be part of the namespace optimization settings (192) of FIG. 10.

[0144] The configuration of FIG. 12 maps a neuron input namespace (185) to at least one storage media component B (163). Preferably, the at least one storage media component B (163) can be used by the controller (151) in parallel with storage media components (e.g., 161) that hosts the other namespaces (e.g., 183 and 187) of ANN data. For example, the storage media component B (163) can be in an integrated circuit package that is separate from the integrated circuit packages for the other namespaces (e.g., 183 and 187). Alternatively, the storage media components (161 to 163) are formed on separate integrated circuit dies embedded in a same integrated circuit package. Alternatively, the storage media components (161 to 163) can be formed on separate regions of an integrated circuit die, where the separate regions can be operated substantially in parallel (e.g., for read, for erase, and/or for write).

[0145] In FIG. 12, the settings (197) are optimized to the usage pattern of enhanced endurance in cyclic sequential overwrite in recording a continuous stream of input data that is sampled at a fixed time interval.

[0146] FIG. 13 illustrates the configuration of a namespace (187) for the outputs from artificial neurons according to one embodiment. For example, the configuration of FIG. 11 can be implemented in the data storage device (101) illustrated in FIGS. 7 and/or 10. For example, the settings (197) of FIG. 11 can be part of the namespace optimization settings (192) of FIG. 10.

[0147] The configuration of FIG. 13 maps a neuron output namespace (187) to at least one storage media component C

(162). Preferably, the at least one storage media component C (162) can be used by the controller (151) in parallel with storage media components (e.g., 161 and 163) that hosts the other namespaces (e.g., 183 and 185) of ANN data. For example, the storage media component C (162) can be in an integrated circuit package that is separate from the integrated circuit packages for the other namespaces (e.g., 183 and 185). Alternatively, the storage media components (161 to 163) are formed on separate integrated circuit dies embedded in a same integrated circuit package. Alternatively, the storage media components (161 to 163) can be formed on separate regions of an integrated circuit die, where the separate regions can be operated substantially in parallel (e.g., for read, for erase, and/or for write).

[0148] In FIG. 13, the settings (197) are optimized to the usage pattern of buffered data for periodic overwrite with random access. For example, memory units are configured via the optimization settings (193 to 197) to update/overwrite in the neuron output namespace (187) at a frequency higher than in the ANN model namespace (183), but lower than in the neuron input namespace (185). For example, the FIG. 14 shows a method of predictive maintenance supported by a model partition according to one embodiment. For example, the method of FIG. 14 can be implemented in a vehicle (111) of FIG. 1 or 2 using a data storage device (101) of FIG. 7 or 10 and/or a storage media component (160) of FIG. 8. For example, the method of FIG. 14 can be used in combination with the method of FIGS. 6 and/or 9.

[0149] At block 241, non-volatile memory of a data storage device (101) is configured into a plurality of partitions (e.g., 183, 185, 187, . . .). For example, the non-volatile memory can have the same type of memory units for storing data (e.g., NAND flash memory units); and the same type of memory units in the different partitions (e.g., 183 to 187) can be configured differently to optimize their performances according to the usage patterns of the data stored in the different partitions (e.g., 183 to 187).

[0150] At block 243, the data storage device (101) stores, for the partitions (e.g., 183, 185, 187, . . .) respectively, different sets of memory operation settings (e.g., 193, 195, 197) for different types of data related to an artificial neural network (125), where the partitions (e.g., 183, 185, 187, . . .) include a model partition (e.g., 193) configured to store model data of the artificial neural network (125).

[0151] At block 245, the data storage device (101) receives a sensor data stream (e.g., 121) from at least one sensor (103) configured on a vehicle (111).

[0152] At block 247, a controller (151) of the data storage device (101) operates memory units in the partitions (183, 185, 187, . . .) according to the sets of memory operation settings (e.g., 193, 195, 197) in response to the sensor data stream (e.g., 121).

[0153] At block 249, the computer system (131) having the data storage device (101) predicts, using the artificial neural network (125), a maintenance service of the vehicle (111) based on the sensor data stream (e.g., 121).

[0154] For example, the memory operation settings configure the model partition (e.g., 183) to store three or more bits per memory cell. The memory operating settings can include address map (191) to map between neurons in the ANN (125) and inputs to the neurons. When a first address of an input to a neuron in the artificial neural network (125) is received, the first address in an input partition (e.g., 185) separate from the model partition (e.g., 183) can be con-

verted into at least one second address of model data associated with the neuron, such that the attributes of the neuron and the identities of neurons connected to the neuron can be retrieved from the model partition (e.g., 183) without an explicit command from the processors (133). The controller (151) can automatically retrieve, from the model partition (e.g., 183), the model data associated with the neuron using the at least one second address, in response to the receiving of the first address. A neural network accelerator (159) can generate an output of the neuron from the input to the neuron and the model data associated with the neuron. In general, input to the neuron can include outputs from multiple neurons that are connected to the neuron in the ANN (125). The controller (151) can save the output of the neuron in the buffer (152) in the data storage device (101) to facilitate accelerated access to the output by the host processor(s) (133) and/or the neural network accelerator (159).

[0155] Typically, the model data does not change during computation to predict the maintenance service. For example, the model data can include neuron connectivity data of the artificial neural network and static attributes of neurons in the artificial neural network. The memory operation settings (e.g., 192) can configure the model partition (e.g., 183) to store more than one bit per memory cell in the non-volatile memory based on the usage pattern of mostly read, infrequent update of the model data.

[0156] For example, the partitions (e.g., 183, 185, 187, . . .) in the data storage devices can be implemented as namespaces in which logical addresses are defined; and an address map (191) in the data storage device is configured to map the namespaces (183, 185, 187, . . .) to separate storage media components (e.g., 161, 163, 162, . . .).

[0157] The model data in the model namespace (183) is updatable during training via machine learning (171 or 175), or during over-the-air update of the ANN (125) from the server (119).

[0158] In some implementations, the controller (151) is configured, via the address map (191) to retrieve, from the model partition, model data associated with a neuron in the artificial neural network, in response to an input to, or an output from, the neuron being addressed in a partition separate from the model partition (183). Further, the controller (151) can retrieve, from the model partition (183), the model data associated with the neuron in parallel with storing input to the neuron in the partition (e.g., 185) that is separate from the model partition (183).

[0159] FIG. 15 shows a method of predictive maintenance supported by an input partition according to one embodiment. For example, the method of FIG. 15 can be implemented in a vehicle (111) of FIG. 1 or 2 using a data storage device (101) of FIG. 7 or 10 and/or a storage media component (160) of FIG. 8. For example, the method of FIG. 15 can be used in combination with the method of FIGS. 6, 9, and/or 14.

[0160] At block 261, non-volatile memory of a data storage device (101) is configured into a plurality of partitions (e.g., 183, 185, 187, . . .). For example, the non-volatile memory can have the same type of memory units (e.g., NAND flash memory units) implemented in multiple storage media components (e.g., 161 to 163).

[0161] At block 263, the data storage device (101) stores, for the partitions (e.g., 183, 185, 187, . . .) respectively, different sets of memory operation settings (e.g., 193, 195, 197) for different types of data related to an artificial neural

network (125), where the partitions (e.g., 183, 185, 187, . . .) include an input partition (e.g., 185) configured to cyclically store input data for the artificial neural network (125).

[0162] For example, the input partition (185) can be configured to store external inputs for the artificial neural network (125) but not internal inputs. The input data stored in the input partition (185) is independent of outputs from neurons in the artificial neural network (125).

[0163] For example, the input data stored in the input partition (185) can include a portion of the sensor data stream (e.g., 121). In some embodiments, the input data stored in the input partition (185) is computed from the sensor data stream (e.g., 121) for a subset of neurons in the artificial neural network (125).

[0164] For example, the memory operation settings (e.g., 195) configure the input partition (185) to store one bit per NAND memory cell in the non-volatile memory for enhanced endurance for repeated data erasure and data programming.

[0165] For example, the memory operation settings (e.g., 195) configure the controller to write sequentially the input data into the input partition (185), and to overwrite oldest input data in the input partition (185) with the most recent input data received in the data storage device (101).

[0166] At block 265, the data storage device (101) receives a sensor data stream (e.g., 121) from at least one sensor (103) configured on a vehicle (111).

[0167] At block 267, a controller (151) of the data storage device (101) operates memory units in the partitions (183, 185, 187, . . .) according to the sets of memory operation settings (e.g., 193, 195, 197) in response to the sensor data stream (e.g., 121).

[0168] At block 269, the computer system (131) having the data storage device (101) predicts, using the artificial neural network (125), a maintenance service of the vehicle (111) based on the sensor data stream (e.g., 121).

[0169] FIG. 16 shows a method of predictive maintenance supported by an input partition according to one embodiment. For example, the method of FIG. 16 can be implemented in a vehicle (111) of FIG. 1 or 2 using a data storage device (101) of FIG. 7 or 10 and/or a storage media component (160) of FIG. 8. For example, the method of FIG. 16 can be used in combination with the method of FIGS. 6, 9, 14, and/or 15.

[0170] At block 281, non-volatile memory of a data storage device (101) is configured into a plurality of partitions (e.g., 183, 185, 187, . . .). The non-volatile memory can have the same type of memory units (e.g., NAND flash memory units) for storing data.

[0171] At block 283, the data storage device (101) stores, for the partitions (e.g., 183, 185, 187, . . .) respectively, different sets of memory operation settings (e.g., 193, 195, 197) for different types of data related to an artificial neural network (125), where the partitions (e.g., 183, 185, 187, . . .) include an output partition (e.g., 187) configured to store output data for the artificial neural network (125).

[0172] For example, the output data stored in the output partition (e.g., 187) can include state data of neurons in the artificial neural network (125). For example, the state data of the neurons in the artificial neural network can identify the activation levels of the neurons for spiking in a spiking neural network. The activation levels can be controlled via a differential equation. Thus, the activation levels can

change in response to inputs to the artificial neural network (125) and/or in response to the passage of time.

[0173] For example, the output data can include the predictions or classifications generated by the artificial neural network (125) responsive to the sensor data stream.

[0174] For example, the memory operation settings configure the output partition to store no more than two bits per memory cell in the non-volatile memory.

[0175] At block 285, the data storage device (101) receives a sensor data stream (e.g., 121) from at least one sensor (103) configured on a vehicle (111).

[0176] At block 287, a controller (151) of the data storage device (101) operates memory units in the partitions (183, 185, 187, . . .) according to the sets of memory operation settings (e.g., 193, 195, 197) in response to the sensor data stream (e.g., 121).

[0177] At block 289, the computer system (131) having the data storage device (101) predicts, using the artificial neural network (125), a maintenance service of the vehicle (111) based on the sensor data stream (e.g., 121).

[0178] For example, the data storage device (101) can include a buffer (152). The buffer (152) can be implemented via volatile memory (e.g., SRAM or DRAM) for access performance faster than the non-volatile memory (e.g., NAND flash memory) of the data storage device (101). The memory operation settings configure the controller (151) to store the output data in the buffer (152) for access by a processor (e.g., 133) via the host interface (157) during or after storing the output data into the output partition (187).

[0179] For example, the data storage device (101) can include a neural network accelerator (159) coupled to the controller (151). The neural network accelerator is configured to apply inputs provided to neurons in the artificial neural network (125) to model data of the artificial neural network (125) to generate the output data by one or more output neurons in the artificial neural network (125). In response to the neural network accelerator (159) completing the computation of the output data, the controller is configured to provide the processor (e.g., 133) with an indication of the availability of the output data generated by the artificial neural network (125), such that the processor (e.g., 133) may request the data storage device (101) to transmit the output data.

[0180] Optionally, the controller (151) is configured to provide the output data to the processor in parallel with storing the output data into the output partition. For example, the controller (151) can be configured to automatically discard the output data computed for the previously segment of sensor data stream if the processor (e.g., 133) does not request for the transmission of the output data to the processor (e.g., 133) within a predetermined period of time, or before the next version of the output data is available. Optionally, after reporting the availability of the output data to the processor (e.g., 133), the controller (151) can be configured to selectively discard the output data computed for the previously segment of sensor data stream based on a response of the processor (e.g., 133) for the output data to the processor (e.g., 133). For example, the processor (e.g., 133) may request the transmission of the output data to the processor (e.g., 133) without saving the output data into the output partition (e.g., 187) in some situations; and in other situations, the processor (e.g., 133) may request the transmission of the output data to the processor (e.g., 133) and the storing of the output data into the output partition (e.g., 187).

[0181] Optionally, output data from the artificial neural network (125) can also be stored into the output partition in a cyclic way (e.g., for the segments of output data for time periods selected by the processor (e.g., 133)).

[0182] For example, external inputs to the artificial neural network (125) can be recorded in the input namespace (185) continuously for the last time period of T1. When the sensor data is sampled at a predetermined time interval T2, the input namespace (185) can hold the latest T1/T2 sets of input data. In contrast, external outputs from the artificial neural network (125) can be selectively recorded into the output namespace (187) (e.g., once for every a predetermined period of time T3, where T3 is multiple of T2). The output data can be recorded into the output namespace (187) at a lower frequency; and the output namespace (187) can be allocated to store a predetermined number of sets of output data (e.g., via sequential writes and writes in a cyclic way to keep the last sets of output data).

[0183] At least some embodiments disclosed herein include a communication protocol/interface that allows a data storage device to perform neural network acceleration on the fly with reduced data traffic to the host processor (e.g., a central processing unit (CPU)).

[0184] For example, the host processor (e.g., 133) of a vehicle (111) can provide write commands to the data storage device (101) to store the model of an artificial neural network in a model partition (e.g., 183). Since the neural network accelerator (159) is configured to apply the model, the data communications sending back the data of the model of ANN (125) to the processor can be reduced or eliminated.

[0185] To use the ANN model in classifications and/or predictions, the host processor (e.g., 133) of a vehicle (111) can stream input data for the ANN (125) into the neuron input partition (e.g., 185). The neural network accelerator (159) of the storage device (101) can automatically apply the input data to the model stored in ANN model partition (e.g., 183) in accordance with the address map (191). The data storage device (101) makes the computed outputs available for propagation in the ANN (125). Preferably, the computed outputs are made available to the neural network accelerator (159) through the buffer (152) without the need to store the intermediate outputs into storage media components (e.g., 161 to 163). Thus, the data communications between the host processor (e.g., 133) and the data storage device (101) for the transporting of outputs of neurons can be reduced. When the outputs have propagated to the output neurons in the ANN (125), the data storage device (101) can provide a response to the write request associating with the writing of a set of input data into the neuron input partition (e.g., 185). The response indicates that the external output from neurons in the ANN (125) is available. In response, the host processor (e.g., 133) of a vehicle (111) can optionally issue read commands to retrieve the external outputs for further processing.

[0186] FIG. 17 shows communications with a data storage device (101) to implement neural network computation according to one embodiment. For example, the communications as illustrated in FIG. 17 can be implemented in the vehicle (111) of FIG. 1 or 2, with a data storage device (101) illustrated in FIG. 7 or 10.

[0187] In FIG. 17, the processor(s) (133) can be configured with a simplified set of instructions (301) to perform neural network computation, since some of the computations involving the ANN (125) is performed by the neural

network accelerator (159) within the data storage device (101). Thus, it is not necessary to transport the model data back to the processor(s) (133) during the use of the ANN (125) for predictions and/or classifications.

[0188] The sensor(s) (103) can generate a continuous stream of sensor data (121) based on a rate for sampling data. The sensor data (121) can be sampled at a fixed, predetermined time interval (e.g., during the operation of the vehicle (111)). The processor(s) (133) can execute the instructions (301) to convert the sensor data (121) into an input stream (303) for input neurons in the ANN (125). Input neurons in the ANN (125) are configured to accept external inputs to the ANN (125); and output neurons are configured to provide external outputs from the ANN (125).

[0189] In general, a complete set of input for the ANN (125) at a time instance includes inputs for the entire set of input neurons of the ANN (125). The input stream (303) includes a sequence of input sets for a sequence of time instances that are spaced apart from each other according to the fixed, predetermined time interval.

[0190] The data storage device (101) stores the input stream (303) into the neuron input namespace (185) in a cyclic way where the oldest input set corresponding to the oldest time instance of data sampling for data sets currently stored in the neuron input namespace (185) is erased to store the newest set of inputs in the input stream (303).

[0191] For each input data set, the neural network accelerator (159) applies the model of the ANN (125) stored in the ANN model namespace (183). The neural network accelerator (159) (or the processor(s) (133)) can control the propagation of signals within the neural network. When the output neurons of the ANN (125) generate their outputs responsive to the input data set, the data storage device (101) can provide to the processor (133) an indication that the neuron output are ready for retrieval. The indication can be configured in a response to the request from the processor(s) (133) to write the input data set into the neuron input namespace (185). The processor(s) (133) can optionally retrieve the output data (305) (e.g., in accordance with conditions and/or criteria programmed in the instructions).

[0192] In some embodiments, a trigger parameter is configured in the data storage device (101). When an output parameter in the external output (317) meetings a requirement specified by the trigger parameter, the data storage device provides the response to the request from the processor(s) (133) to write the input data set into the neuron input namespace (185).

[0193] FIG. 18 shows communications within a data storage device to implement neural network computation according to one embodiment. For example, the communications of FIG. 18 can be implemented a data storage device (101) illustrated in FIG. 7 or 10, in connection with the communications of FIG. 17.

[0194] In FIG. 18, the model namespace (183) stores the model (313) of the entire ANN (125). In response to receiving a set of external input (315) for a time instance from the input stream (303) in the buffer (152), the data storage device (101) can write the external input (315) into the input namespace (185) in parallel with retrieving a neuron model (312) containing a portion of the ANN model (313) corresponding to the parameters of the input neurons and/or the identities of neurons connected to the input neurons. The buffer (152) allows the neural network accel-

erator (159) to combine the neuron model (312) and the external input (325) to generate the output (327) of the input neurons.

[0195] In general, the neuron output (327) can include a portion that is the internal output (316) for further propagation within the ANN (125) and/or a portion that is the external output (317) for the processor(s) (133).

[0196] The internal output (316) is stored in the buffer (152) as internal input (316) for further propagation in the ANN (125) in a way similar to the generation of neuron outputs (327) from the external input (315). For example, a portion of the internal input (316) can cause the controller (151) and/or the neural network accelerator (159) to retrieve corresponding neuron model (312) relevant to the internal input such that the internal input is applied in the neural network accelerator (159) to the corresponding neuron model (312) to generate their neuron outputs (327).

[0197] When the complete set of external output (317) is available in the buffer (152), the external output (317) can be stored in the output namespace (187).

[0198] Optionally, the storage device (101) does not store each set of external output (317) corresponding to a set of stored external input (315) sampled at a time instance. For example, the storage device (101) can be configured to store one set of external output (317) for every a predetermined number of sets of external input (e.g., 315). Alternatively, or in combination, the processor(s) (133) can determine whether or not to store the external output (317). For example, the storage device (101) can be configured to store the external output (317) in response to the processor(s) (133) retrieving the external output (317) for further processing. For example, the storage device (101) can be configured to store the external output (317) in response to a write command from the processor(s) (133) after the processing of the external output (317) in the processor(s) (133).

[0199] FIG. 19 shows a method of communicating with a data storage device to implement neural network computation according to one embodiment. For example, the method of FIG. 19 can be implemented in a vehicle (111) of FIG. 1 or 2 using a data storage device (101) of FIG. 7 or 10 and/or a storage media component (160) of FIG. 8. For example, the method of FIG. 19 can be used in combination with the method of FIGS. 6, 9, 14, 15, and/or 16.

[0200] At block 341, one or more processors (133) of a vehicle (111) stores model data (e.g., 313) of an artificial neural network (e.g., 125) into a data storage device (101).

[0201] At block 343, the one or more processors (133) of the vehicle (111) receive, from at least one sensor (103) configured on vehicle (111), a set of sensor data.

[0202] At block 345, the one or more processors (133) of the vehicle (111) generate a set of inputs to the artificial neural network (e.g., 125) based on the sensor data.

[0203] At block 347, the one or more processors (133) of the vehicle (111) provide the set of inputs to the data storage device (101). In response to the set of inputs, the data storage device (101) is configured to generate a set of outputs using the model data (313) of the artificial neural network (e.g., 125).

[0204] At block 349, the one or more processors (133) of the vehicle (111) retrieve the set of outputs from the data storage device (101).

[0205] For example, the data storage device (101) generates the set of outputs using at least a portion of the model

data (183) stored in the data storage device without transmitting the portion of the model data (183) to the one or more processors (133) between the receiving of the set of inputs and the completion of the computation of the set of outputs.

[0206] For example, the portion of the model data (183) can include static attributes of neurons in the artificial neural network (e.g., 125) and/or the neuron connectivity data of the artificial neural network (e.g., 125).

[0207] For example, to provide of the set of inputs to the data storage device (101), the one or more processors (133) of the vehicle (111) can transmit one or more write commands to the data storage device (101). The one or more write commands are configured to instruct the data storage device (101) to store the set of input in the data storage device (101). After the completion of the computation of the set of outputs in the data storage device (101), the controller (151) of the data storage device (101) can transmit a response to the one or more write commands to the one or more processors (133). The response can include an indication that the set of outputs is available for retrieval by the one or more processors (133).

[0208] In response to the indication, the one or more processors (133) can optionally retrieve of the set of outputs from the data storage device (101) by transmitting a read command to the data storage device (101) for the set of outputs (e.g., after a determination to retrieve the set of outputs from the data storage device (101) for processing).

[0209] Alternatively, or in combination, the one or more processors (133) of the vehicle (111) can determine whether to store the set of outputs in non-volatile memory of the data storage device. In response to a determination to store the set of outputs in the non-volatile memory of the data storage device (101), one or more processors (133) of the vehicle (111) can transmit a write command to the data storage device (101).

[0210] Since the set of outputs is initially generated in the data storage device (101) and then buffered in the buffer (152) (e.g., volatile memory), the data storage device (101) can execute a write command to store the set of outputs into an output namespace (187) without transmitting the set of outputs to the one or more processors (133) and/or receiving the set of outputs from the one or more processors (133) in response to the write commands.

[0211] For example, after receiving, from at least one sensor (103) configured on the vehicle (111), a further set of sensor data (121), the one or more processors (133) of the vehicle (111) generate a further set of inputs to the artificial neural network (125) based on the further set of sensor data.

[0212] The one or more processors (133) transmits a further command to write the further set of inputs into the data storage device (101); and the data storage device (101) generates a further set of outputs using the model data (183) of the artificial neural network (125) and the further set of inputs. After receiving a response to the further command to write the further set of inputs, the one or more processors (133) can determine to skip the processing of the further set of outputs and transmit, to the data storage device (101), a subsequent write command to store the further set of outputs. In response, the data storage device (101) can write the further set of outputs that is buffered within the data storage device (101) into the output namespace (187), without the transmitting of the further set of outputs from the one or more processors (133) of the vehicle (111) to the data storage

device (101) and/or without the transmitting of the further set of outputs from the data storage device (101) to the one or more processors (133) of the vehicle (111).

[0213] The techniques discussed above can be used in predictive maintenance of automotive battery.

[0214] For example, a device with limited processing power and storage capacity can be used to monitor the operating conditions of battery and/or the system power for a most recent period of time. Based on the monitored parameters as a function of time, an artificial neural network (125) (or a statistical model) can predict whether a maintenance service is recommended for the battery within a period of time. Optionally, the artificial neural network (125) (or the statistical model) is configured to predict a time period to the occurrence of a battery failure (e.g., insufficient power or voltage) if no maintenance service is provided within the time period.

[0215] Optionally, such a device can be built in the battery pack as a component of the battery. Alternatively, a portion of the processing resource and storage capacity of a data storage device (101) can be configured as the device for monitoring the health of the battery.

[0216] An artificial neural network (125) in the battery-monitoring device can be trained to predict a time period to a recommended maintenance. When the predicted time period is less than a threshold, the device can generate an alert in the infotainment system (e.g., 149) of the vehicle (111) and/or the on-board vehicle computer (e.g., 131) of the vehicle (111).

[0217] For example, the battery in the vehicle (111) can be configured to power an engine starter that rotates or cranks the internal combustion engine of the vehicle (111) during the initial operations of the engine. When the battery does not have sufficient voltage or stored energy, the engine starter is not capable of assisting the internal combustion engine in starting its operations. The battery is recharged after the initiation of the operations of the internal combustion engine. Over many cycles of discharge and recharge, the performance capacity of the battery may degrade such that the internal combustion engine cannot sufficiently recharge the battery to have sufficient voltage or stored energy for the operation of the engine starter. Therefore, when the performance capacity of the battery is sufficiently degraded, a battery service is recommended to avoid an incident where the internal combustion engine of the vehicle (111) cannot be started.

[0218] The battery configured to power an engine starter can also be used to power accessories of the vehicle (111), such as the infotainment system (149) and/or the computer system (131) of the vehicle (111), when the internal combustion engine of the vehicle (111) is not running. However, the usage of the accessories for an extended period powered by the battery can sufficiently drain the battery to prevent a successful start of the internal combustion engine of the vehicle (111). The battery-monitoring device can be configured to monitor the state of the battery to predict when the battery is close to a critical state such that when the battery is further drained, the battery is likely to have insufficient power for the engine starter to initiate the operations of the internal combustion engine. When the battery is predicted to be close to the critical state, recharging the battery (e.g., via starting the internal combustion engine) and/or reducing/stopping the use of the accessories can be recommended via

alerts presented in the infotainment system (149) and/or the computer system (131) of the vehicle (111).

[0219] The battery configured to power an engine starter can have problems such as corrosion in terminals, loosen connections, low electrolyte levels, etc. The artificial neural network (125) in the battery-monitoring device can be trained to predict, based on the monitored parameters of the battery, whether the battery has any of the problems.

[0220] In some configurations, the battery of the vehicle (111) is configured to power one or more electric motors that drive the vehicle (111). The battery-monitoring device can determine whether the battery of the vehicle (111) is sufficiently degraded to reduce the driving range of the vehicle (111). When the reduction of the driving range of the vehicle (111) is above a threshold, a maintenance service recommendation can be presented via the infotainment system (149) and/or the computer system (131) of the vehicle (111).

[0221] Optionally, when the battery is operating in a time period in which the battery can be assumed to be operating in a normal condition, the artificial neural network (125) can be trained to recognize the normal patterns of battery operating parameters. Subsequently, when the pattern of battery operating parameters deviates from the normal patterns, the battery-monitoring device can suggest a battery maintenance service.

[0222] Data of abnormal battery operating parameters can be stored in the data storage device (101) of the vehicle (111) automatically for subsequent diagnosis in a maintenance facility (127). After the inspection and/or diagnosis in the maintenance facility (127), the inspection and/or diagnosis result generated in the maintenance facility (127) can be used to further train the artificial neural network (125) in the server (119) to improve the predictive capability of the artificial neural network (125). The improved artificial neural network (125) can be downloaded to the vehicle (111) and/or other vehicles having similar battery to improve predictive maintenance services.

[0223] Data of normal battery operating parameters can be discarded by the data storage device (101) to reduce the need for a large storage capacity.

[0224] For example, the artificial neural network (125) can be initially trained to at least recognize the patterns of operating parameters of the battery of the vehicle (111) when the battery is considered to be operating in a normal working condition but in the environment that is specific to the vehicle (111). The battery operating parameters indicate the different patterns of battery usage, such as power usage by accessories and/or add-on devices connected to the battery. The training can reduce the interference of the noises from the accessories and/or add-on devices on the predictive capability of the artificial neural network (125) in recognizing abnormal operating characteristics in the monitored operating parameters of the battery.

[0225] For example, when the vehicle (111) has less than a predetermined of miles from leaving a factory and/or a maintenance service facility (127) after the installation of a new battery, and/or has less than a predetermined of times of starting the engine of the vehicle (111), the battery of the vehicle (111) can be assumed to be operating normally. Thus, the artificial neural network (125) can be trained to recognize the additional noises that are specific to the daily/routine operating environment of the vehicle (111) but are not indication of anomaly. Subsequently, deviation from

the patterns of normal operations can be detected by the artificial neural network (125) for a classification of anomaly.

[0226] Data of abnormal battery operations can be collected in the maintenance facility (127), which can further analyze the data of abnormal battery operation data and/or inspect the battery to label the data patterns with diagnosis. Abnormal battery operation data with diagnosis from a population of similar vehicles can be used in the server (119) to train the ANN (125) to predict the diagnosis associated with the abnormal battery operations. Once the artificial neural network (125) is updated to include the capabilities to predict the diagnosis, predictive maintenance suggestions generated by the vehicle (111) can be more precise.

[0227] FIG. 20 shows the use of an artificial neural network (ANN) to monitor the health of automotive battery according to one embodiment. For example, the neural network computation illustrated in FIG. 20 can be implemented with a data storage device (101) of FIG. 7 or 10 for a battery in a vehicle (111) illustrated in FIG. 1 or 2.

[0228] In FIG. 20, the battery (351) provides power to one or more components of the vehicle (111), such as an electric motor of an electric vehicle (111) or an engine starter of a vehicle having an internal combustion engine, a fan, an air conditioning unit, and/or electric or electrical accessories configured in the vehicle (111).

[0229] In FIG. 20, a current sensor (353) and a voltage sensor (355) are configured to monitor the operating conditions of the battery (351). Addition sensors, such as a temperature sensor, can be used to monitor the operating conditions of the battery (351). The sensors (e.g., 353 and 355) configured to monitor the operating conditions of the battery (351) can be part of the sensors (103) of the vehicle (111).

[0230] The operating parameters measured by the sensors (353 and 355) as a function of time can be provided as a sensor data stream to the artificial neural network (125) to generate predictions/classifications (359).

[0231] In general, the operating parameters generated by the sensors (353 and 355) reflect not only the characteristics of the battery (351) but also the loads powered by the battery (351), such as the engine starter, a fan, an air conditioning unit, and/or electric or electrical accessories configured in the vehicle (111). The ANN (125) can be trained to recognize the patterns of normal operations of different combinations of the loads of the battery (351) such that an operation of a load or accessory specific to the vehicle (111) in which the battery (351) is configured does not trigger a false alarm and/or interfere with the classification or prediction results of the artificial neural network (125).

[0232] The sensors (353 and 355) can be configured on the battery (351). Alternatively, a set of the sensors (353 and 355) can be configured at a centralized location on the vehicle (111) and dynamically connected to the battery for a time period to monitor the operation of the battery (351). For example, the vehicle (111) can have multiple battery packs; and the set of the sensors (353 and 355) can be switched to monitor any selected one of the battery packs.

[0233] Optionally, the artificial neural network (ANN) (125) can be trained using an unsupervised learning technique as illustrated in FIG. 4, or a supervised learning technique as illustrated in FIG. 5, to recognize the load

noises generated by the operations of a fan, an air conditioning unit, and/or electric or electrical accessories configured in the vehicle (111).

[0234] The artificial neural network (ANN) (125) is configured to generate classifications and/or predictions (359) based on the measurements generated by the sensors (353 and 355).

[0235] For example, the artificial neural network (ANN) (125) can classify whether the characteristics of the current-voltage curve during a time period of discharge or recharge, as measured by the sensors (353 and 355), are normal or abnormal. When the characteristics of the current-voltage curve during discharge or recharge is abnormal, the data storage device (101) stores a copy of the abnormal current-voltage curve; and the processors (133) of the vehicle (111) can generate an alert on the infotainment system (149) to suggest a schedule for a maintenance service.

[0236] Optionally, the communication device (139) is controlled by the processors (133) to communicate with the maintenance service facility (127) to obtain time slots for the maintenance service.

[0237] Optionally, the communication device (139) is controlled by the processors (133) to provide the copy of the abnormal current-voltage curve to the maintenance service facility (127) for analysis and/or for a suggestion for maintenance services.

[0238] For example, the artificial neural network (ANN) (125) can be trained to recognize/predict the characteristics of the current-voltage curve of some common battery problems. When a battery problem is predicted, the vehicle (111) can be configured to prompt the user of the vehicle (111) (e.g., a driver or passenger) to schedule a visit to the maintenance service facility (127).

[0239] In some embodiments, the artificial neural network (ANN) (125) receives not only the current-voltage measurements generated by the sensors (353 and 355) but also the operation signals of the vehicle (111). The vehicle operation signals are indicative of the work condition and/or the workload of the battery (351). For example, the vehicle operation signals can indicate whether the engine of the vehicle (111) is starting, idling, accelerating, braking, etc. For example, the vehicle operation signals can indicate whether a fan, an air conditioning unit, and/or electric or electrical accessories configured in the vehicle (111) are turned on and being powered by the battery (351). For example, the vehicle operation signals can include the signals related to the statuses of the acceleration control (145), the braking control (143), the steering control (141), the speed of the vehicle (111), etc. The correlation between the detected load noises and the vehicle operation signals can improve the accuracy in the predictions/classifications (359) generated by the artificial neural network (ANN) (125).

[0240] Optionally, the vehicle operation signals are applied in the artificial neural network (ANN) (125) to classify the load state of the battery (351); and the artificial neural network (ANN) (125) is further configured to produce the predictions/classifications (359) based on the sensor measurements corresponding to one or more pre-selected load state of the battery (351).

[0241] Alternatively, the instructions (301) for the processors (133) are configured to determine the current load state of the battery (351); and the processors (133) are configured to selectively use the sensors (353 and 355) to generate

measurements based on the load state of the battery (351). The current voltage curve selectively recorded according to the current load state of the battery (351) and/or the schedule for monitoring can be provided to the data storage device (101) (e.g., in a way as illustrated in FIG. 17) to generate the predictions/classifications (359). The selective recording and/or analysis of the battery sensor data (e.g., from sensors 353 and 355) can reduce the data communications to the data storage device (101).

[0242] FIG. 21 shows a method of predictive maintenance of battery of a vehicle according to one embodiment. For example, the method of FIG. 21 can be implemented in a vehicle (111) of FIG. 1 or 2 using a data storage device (101) of FIG. 7 or 10 and/or a storage media component (160) of FIG. 8. For example, the method of FIG. 21 can be used in combination with the method of FIGS. 6, 9, 14, 15, 16, and/or 19.

[0243] At block 361, one or more sensors (e.g., 353 and/or 355) measures operating parameters of battery (351) of a vehicle (111). The battery sensor(s) (e.g., 353 and/or 355) can be part of the sensors (103) of the vehicle (111). The battery sensor(s) (e.g., 353 and/or 355) can be configured as part of the battery (351), or be connected to the battery (351) from a centralized location on the vehicle (111).

[0244] At block 363, the operating parameters of the battery (351) are provided as a function of time to an artificial neural network (125) as input.

[0245] For example, the battery (351) is configured primarily to provide power for an electric motor of the vehicle, such as the motor of an engine starter for a vehicle having an internal combustion engine, or a drive motor of an electric vehicle that does not have an internal combustion engine (or having an internal combustion engine configured as a power generator for the drive motor).

[0246] The battery (351) can also be configured to provide power to secondary devices, such as a fan, air conditioning unit, motors for power windows, a motor for windshield wipers, a radio, an USB charging port, a cigarette lighter, etc.

[0247] At block 365, the operating parameters of the battery (351) as a function of time are analyzed via the artificial neural network (125) to generate a result.

[0248] At block 367, the vehicle (111) generates a suggestion for a maintenance service of the battery (351) based on the result from the artificial neural network (125) analyzing the operating parameters.

[0249] For example, the result from the ANN (125) can include an identification of a battery problem determined from the operating parameters measured by the sensor(s) (353 and/or 355), and/or a classification of whether the operations of the battery (351) are normal.

[0250] For example, the artificial neural network (125) can be initially trained to recognize patterns of normal loads of the battery (351) during a time period in which operations of the battery (351) is considered to be normal. Such a time period can be a predetermined time period from a new battery (351) being installed/or tested (e.g., in a factory, or in a maintenance service facility (127)), or a predetermined time period after a routine maintenance service.

[0251] For example, the artificial neural network (125) can be a spiking neural network; and the initial training in the time period can be performed using an unsupervised machine learning technique to recognize the load noises in typical usages of the battery (351) that can be specific to the

usage environment of the specific vehicle (111). After the training, deviation from the normal load patterns can be detected to suggest a maintenance service of the battery (351).

[0252] For example, the vehicle can have a data storage device (101) configured to store model data (313) of the artificial neural network (125) and compute the classification/prediction result of the ANN (125) based on the model data (313) stored in the data storage device (101) and based on the inputs to the artificial neural network (125) received from the processor(s) (133) of the vehicle (111). For example, a neural network accelerator (159) of the data storage device (101) can apply the inputs to the artificial neural network (125) without transmitting the model data (313) from the data storage device (101) to the processor(s) (133).

[0253] The inputs to the ANN (125) can be data representing the current voltage curves measured by the battery sensor(s) (e.g., 353 and 355) and/or load features extracted by the processor(s) (133) executing instructions (301).

[0254] Optionally, the processor(s) (133) can provide not only the battery operating parameters measured by the battery sensor(s) (e.g., 353 and 355), to the data storage device (101) as input to the ANN (125), but also the operation signals of the vehicle (111) as inputs to the ANN (125) at corresponding time instances of the battery operating parameters. Alternatively, the processor(s) (133) and/or the ANN (125) can analyze the vehicle operation signals to determine the current load state of the battery (351) and selectively transmit the battery operating parameters for the ANN (125) based on the current load state of the battery (351).

[0255] The data storage device (101) is configured to store the data representing operating parameters of the battery (351) when the result of the ANN (125) indicates abnormal battery operations and/or transmit the data representing the battery operating parameters to a maintenance service facility (127). Optionally, the vehicle (111) can communicate with the maintenance service facility (127) to schedule a trip for the suggested maintenance service and/or present the suggestion via the infotainment system (149) of the vehicle (111).

[0256] Optionally, the data of abnormal battery operating parameters stored in the data storage device (101) can be downloaded to the maintenance service facility (127) during the performance of the maintenance service of the battery (351). The diagnosis of the battery (351) having the abnormal operating parameters can be transmitted with the data of the abnormal operating parameters to a server (119). A supervised machine learning technique can be applied to the ANN (125) to train the ANN (125) in predicting, based on the data of the abnormal operating parameters, the diagnosis of the battery (351) having the abnormal operating parameters.

[0257] The server (119), the computer system (131), and/or, the data storage device (101) can each be implemented as one or more data processing systems.

[0258] The present disclosure includes methods and apparatuses which perform the methods described above, including data processing systems which perform these methods, and computer readable media containing instructions which when executed on data processing systems cause the systems to perform these methods.

[0259] A typical data processing system may include includes an inter-connect (e.g., bus and system core logic), which interconnects a microprocessor(s) and memory. The microprocessor is typically coupled to cache memory.

[0260] The inter-connect interconnects the microprocessor (s) and the memory together and also interconnects them to input/output (I/O) device(s) via I/O controller(s). I/O devices may include a display device and/or peripheral devices, such as mice, keyboards, modems, network interfaces, printers, scanners, video cameras and other devices known in the art. In one embodiment, when the data processing system is a server system, some of the I/O devices, such as printers, scanners, mice, and/or keyboards, are optional.

[0261] The inter-connect can include one or more buses connected to one another through various bridges, controllers and/or adapters. In one embodiment the I/O controllers include a USB (Universal Serial Bus) adapter for controlling USB peripherals, and/or an IEEE-1394 bus adapter for controlling IEEE-1394 peripherals.

[0262] The memory may include one or more of: ROM (Read Only Memory), volatile RAM (Random Access Memory), and non-volatile memory, such as hard drive, flash memory, etc.

[0263] Volatile RAM is typically implemented as dynamic RAM (DRAM) which requires power continually in order to refresh or maintain the data in the memory. Non-volatile memory is typically a magnetic hard drive, a magnetic optical drive, an optical drive (e.g., a DVD RAM), or other type of memory system which maintains data even after power is removed from the system. The non-volatile memory may also be a random access memory.

[0264] The non-volatile memory can be a local device coupled directly to the rest of the components in the data processing system. A non-volatile memory that is remote from the system, such as a network storage device coupled to the data processing system through a network interface such as a modem or Ethernet interface, can also be used.

[0265] In the present disclosure, some functions and operations are described as being performed by or caused by software code to simplify description. However, such expressions are also used to specify that the functions result from execution of the code/instructions by a processor, such as a microprocessor.

[0266] Alternatively, or in combination, the functions and operations as described here can be implemented using special purpose circuitry, with or without software instructions, such as using Application-Specific Integrated Circuit (ASIC) or Field-Programmable Gate Array (FPGA). Embodiments can be implemented using hardwired circuitry without software instructions, or in combination with software instructions. Thus, the techniques are limited neither to any specific combination of hardware circuitry and software, nor to any particular source for the instructions executed by the data processing system.

[0267] While one embodiment can be implemented in fully functioning computers and computer systems, various embodiments are capable of being distributed as a computing product in a variety of forms and are capable of being applied regardless of the particular type of machine or computer-readable media used to actually effect the distribution.

[0268] At least some aspects disclosed can be embodied, at least in part, in software. That is, the techniques may be

carried out in a computer system or other data processing system in response to its processor, such as a microprocessor, executing sequences of instructions contained in a memory, such as ROM, volatile RAM, non-volatile memory, cache or a remote storage device.

[0269] Routines executed to implement the embodiments may be implemented as part of an operating system or a specific application, component, program, object, module or sequence of instructions referred to as "computer programs." The computer programs typically include one or more instructions set at various times in various memory and storage devices in a computer, and that, when read and executed by one or more processors in a computer, cause the computer to perform operations necessary to execute elements involving the various aspects.

[0270] A machine readable medium can be used to store software and data which when executed by a data processing system causes the system to perform various methods. The executable software and data may be stored in various places including for example ROM, volatile RAM, non-volatile memory and/or cache. Portions of this software and/or data may be stored in any one of these storage devices. Further, the data and instructions can be obtained from centralized servers or peer to peer networks. Different portions of the data and instructions can be obtained from different centralized servers and/or peer to peer networks at different times and in different communication sessions or in a same communication session. The data and instructions can be obtained in entirety prior to the execution of the applications. Alternatively, portions of the data and instructions can be obtained dynamically, just in time, when needed for execution. Thus, it is not required that the data and instructions be on a machine readable medium in entirety at a particular instance of time.

[0271] Examples of computer-readable media include but are not limited to non-transitory, recordable and non-recordable type media such as volatile and non-volatile memory devices, read only memory (ROM), random access memory (RAM), flash memory devices, floppy and other removable disks, magnetic disk storage media, optical storage media (e.g., Compact Disk Read-Only Memory (CD ROM), Digital Versatile Disks (DVDs), etc.), among others. The computer-readable media may store the instructions.

[0272] The instructions may also be embodied in digital and analog communication links for electrical, optical, acoustical or other forms of propagated signals, such as carrier waves, infrared signals, digital signals, etc. However, propagated signals, such as carrier waves, infrared signals, digital signals, etc. are not tangible machine readable medium and are not configured to store instructions.

[0273] In general, a machine readable medium includes any mechanism that provides (i.e., stores and/or transmits) information in a form accessible by a machine (e.g., a computer, network device, personal digital assistant, manufacturing tool, any device with a set of one or more processors, etc.).

[0274] In various embodiments, hardwired circuitry may be used in combination with software instructions to implement the techniques. Thus, the techniques are neither limited to any specific combination of hardware circuitry and software nor to any particular source for the instructions executed by the data processing system.

[0275] The above description and drawings are illustrative and are not to be construed as limiting. Numerous specific

details are described to provide a thorough understanding. However, in certain instances, well known or conventional details are not described in order to avoid obscuring the description. References to one or an embodiment in the present disclosure are not necessarily references to the same embodiment; and, such references mean at least one.

[0276] In the foregoing specification, the disclosure has been described with reference to specific exemplary embodiments thereof. It will be evident that various modifications may be made thereto without departing from the broader spirit and scope as set forth in the following claims. The specification and drawings are, accordingly, to be regarded in an illustrative sense rather than a restrictive sense.

What is claimed is:

1. A vehicle, comprising:
 - an electric motor;
 - battery configured to power at least the electric motor;
 - one or more sensors configured to measure operating parameters of the battery;
 - an artificial neural network configured to analyze the operating parameters of the battery as a function of time to generate a result; and
 - at least one processor configured to generate a suggestion for a maintenance service of the battery based on the result from the artificial neural network analyzing the operating parameters of the battery.
2. The vehicle of claim 1, wherein the result includes an identification of a battery problem determined from the operating parameters measured by the one or more sensors.
3. The vehicle of claim 1, wherein the result includes a classification of whether battery operations are normal.
4. The vehicle of claim 3, wherein the artificial neural network is trained to recognize patterns of loads of the battery during a time period in which battery operations of the vehicle is considered to be normal.
5. The vehicle of claim 4, wherein the artificial neural network is a spiking neural network.
6. The vehicle of claim 1, further comprising:
 - a data storage device configured to store model data of the artificial neural network and compute the result based on the model data stored in the data storage device and inputs to the artificial neural network received from the at least one processor.
7. The vehicle of claim 6, wherein the at least one processor is configured to generate the inputs based on measurements generated by the one or more sensors.
8. The vehicle of claim 6, wherein the one or more sensors include a current sensor and a voltage sensor.
9. The vehicle of claim 8, wherein the inputs to the artificial neural network for generation of the result further include operation signals of the vehicle.
10. The vehicle of claim 9, wherein the data storage device is configured to store data representing current-voltage curve during discharge or recharge of the battery when the result indicates abnormal battery operations.

11. A method, comprising:
 - measuring, by one or more sensors, operating parameters of battery of a vehicle;
 - providing the operating parameters of the battery as a function of time to an artificial neural network;
 - analyzing, via the artificial neural network, the operating parameters of the battery as a function of time to generate a result; and
 - generating a suggestion for a maintenance service of the battery based on the result from the artificial neural network analyzing the operating parameters.
12. The method of claim 11, wherein the result includes an identification of a battery problem determined from a current voltage curve of the battery identified in the operating parameters, or a classification of whether operations of the battery are normal.
13. The method of claim 11, further comprising:
 - training, in the vehicle, the artificial neural network to recognize load noises for the battery during a time period in which operations of the battery is predetermined to be normal.
14. The method of claim 11, further comprising:
 - presenting the suggestion in an infotainment system of the vehicle.
15. The method of claim 11, further comprising:
 - in response to a classification that operations of the battery are abnormal, transmitting the operating parameters to a maintenance service facility.
16. The method of claim 15, further comprising:
 - communicating with the maintenance service facility to schedule a trip for the maintenance service, in response to the suggestion.
17. The method of claim 11, further comprising:
 - in response to a classification that operations of the battery are abnormal, storing the data representing the operating parameters in non-volatile memory of a data storage device configured on the vehicle; and
 - providing a maintenance service facility with the data representing the operating parameters during the maintenance service.
18. The method of claim 17, further comprising:
 - training the artificial neural network to predict, based on the data representing operating parameters, a battery problem diagnosed in the maintenance service.
19. A battery pack, comprising:
 - battery configured to power an electric motor of a vehicle;
 - one or more sensors configured to measure operating parameters of the battery;
 - an artificial neural network configured to analyze the operating parameters of the battery as a function of time to generate a result; and
 - a processor configured to generate a suggestion for a maintenance service of the battery based on the result from the artificial neural network analyzing the operating parameters of the battery.
20. The battery pack of claim 19, wherein the one or more sensors include a current sensor and a voltage sensor.

* * * * *