

(12) **UK Patent Application** (19) **GB** (11) **2582124** (13) **A**

(43) Date of A Publication **16.09.2020**

(21) Application No: **1814620.9**

(22) Date of Filing: **07.09.2018**

(51) INT CL:  
**A61B 5/00** (2006.01) **G06K 9/62** (2006.01)  
**G16H 50/20** (2018.01)

(71) Applicant(s):  
**Transformative AI Ltd**  
**14 Carlton Court, Carlton Way, Cambridge,**  
**Cambridgeshire, CB4 2BX, United Kingdom**

(56) Documents Cited:  
**GB 2560339 A** **EP 3120753 A1**  
**WO 2016/183683 A1** **US 20170196769 A1**  
**US 20050137482 A1** **US 20020068875 A1**

(72) Inventor(s):  
**Marek Sirendi**  
**Joshua Oppenheimer**  
**Marek Rei**

(58) Field of Search:  
 INT CL **A61B**  
 Other: **EPODOC, WPI**

(74) Agent and/or Address for Service:  
**Mathys & Squire LLP**  
**The Shard, 32 London Bridge Street, LONDON,**  
**SE1 9SG, United Kingdom**

(54) Title of the Invention: **Analysis of cardiac data**  
 Abstract Title: **Analysing cardiac data to indicate probability of cardiac event**

(57) A method of analysing cardiac data provided for example by an electrocardiogram (ECG), cardioverter defibrillator or capnography, comprises determining a property of the data e.g. a mean RR interval, determined over a particular context length selected based on the property, for instance around 230 beats. This is compared with a threshold to indicate a probability of the patient experiencing a cardiac event within a risk window, which may be a probability density distribution. The threshold may be determined by training classifiers to classify properties of multiple heartbeats using a machine learning algorithm. The method may enable preventative action to be taken, or may enable determination of periods during which increased patient monitoring would be beneficial. Also disclosed is a method of training a hybrid classifier by combining classifiers based upon a performance metric, such as accuracy, sensitivity, specificity or precision.

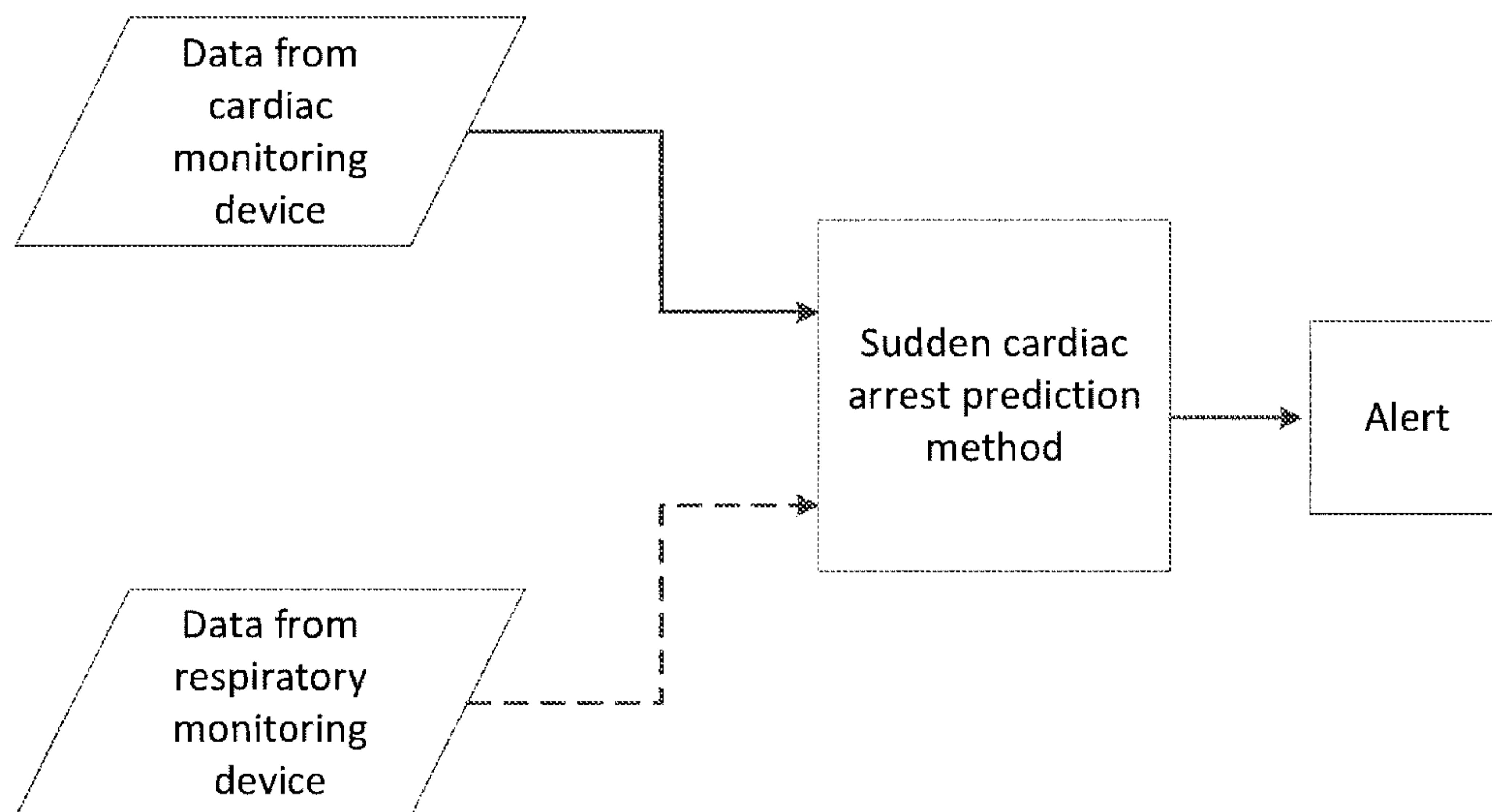


Figure 2a

GB 2582124 A

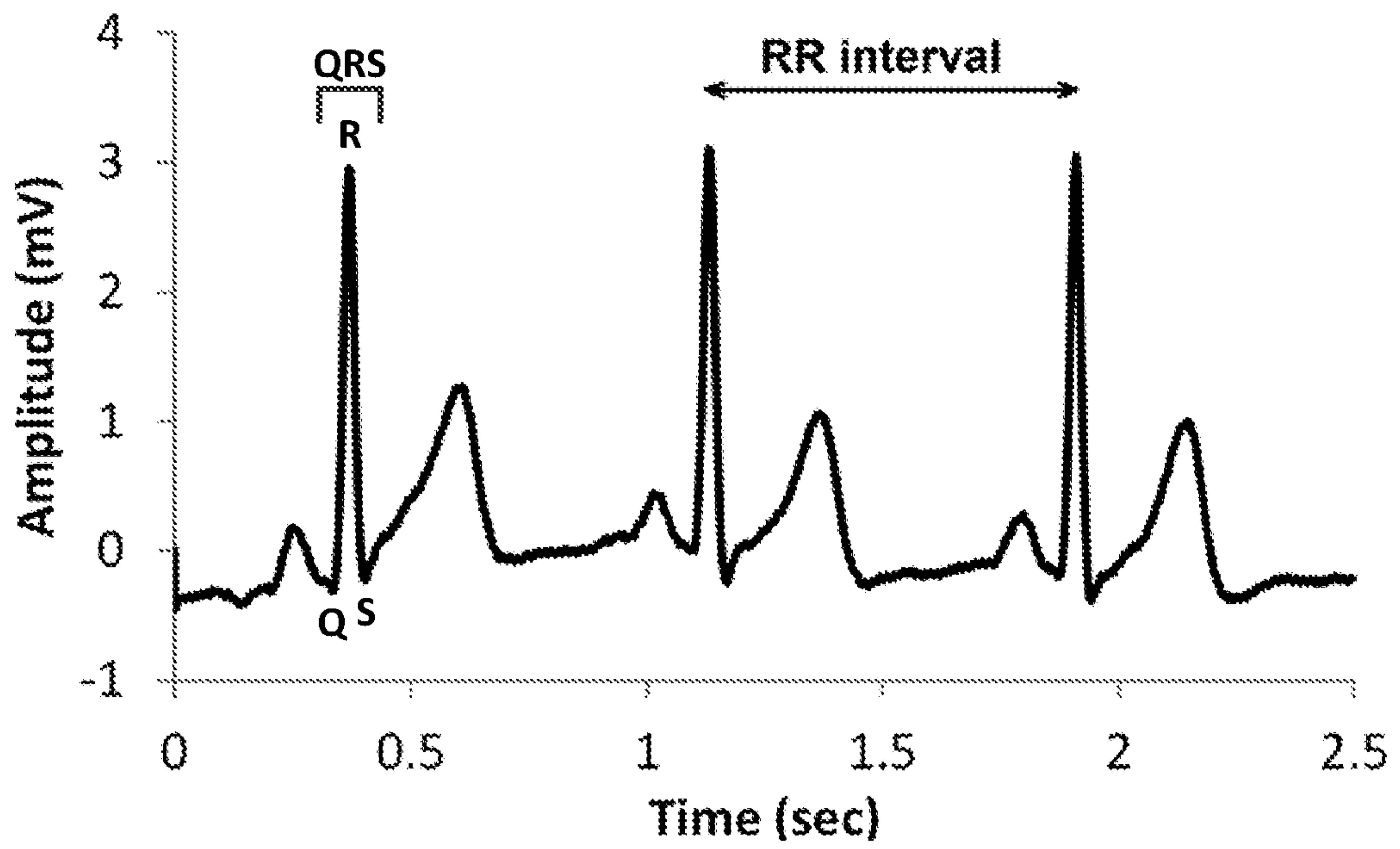


Figure 1

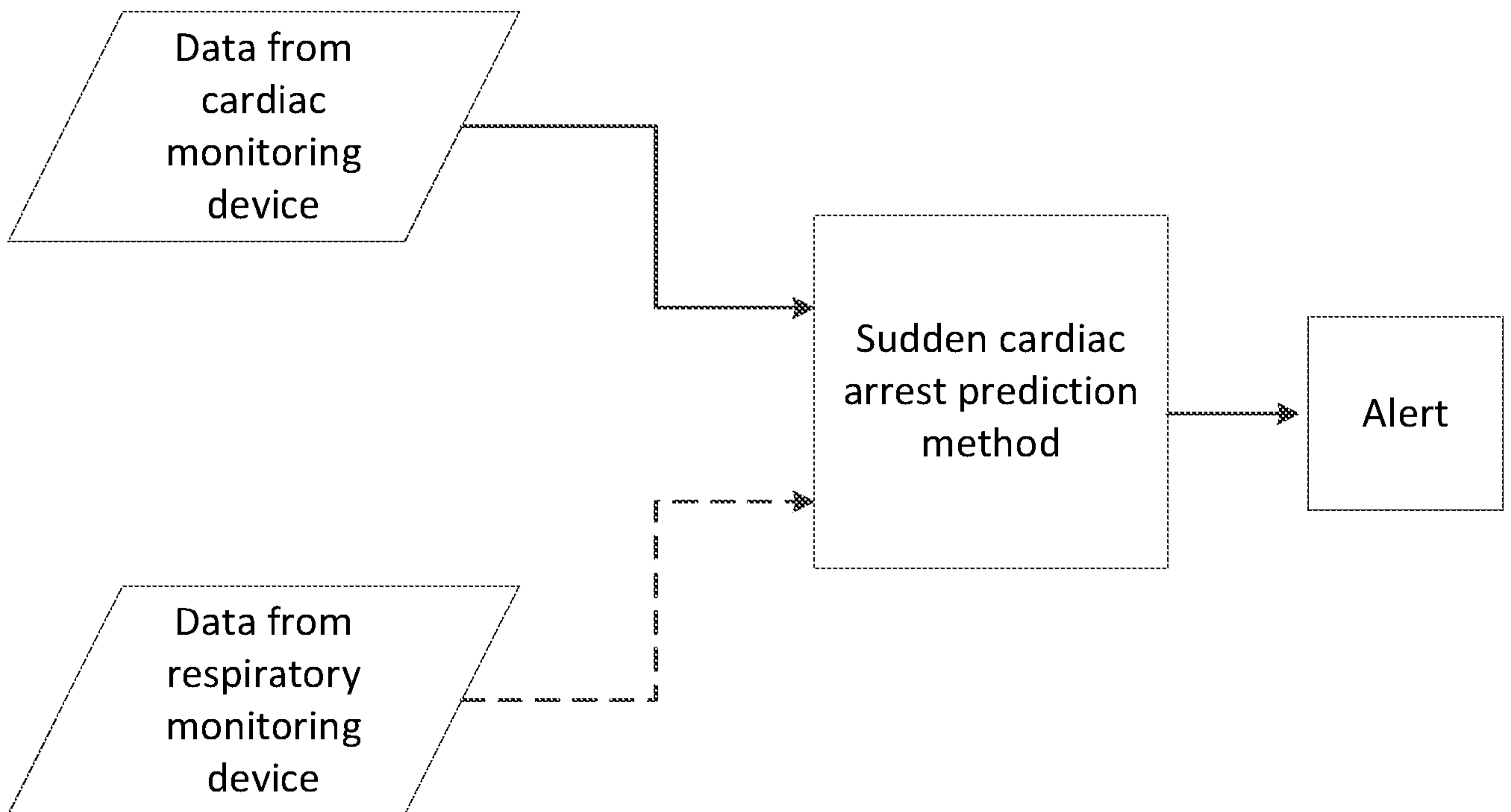


Figure 2a

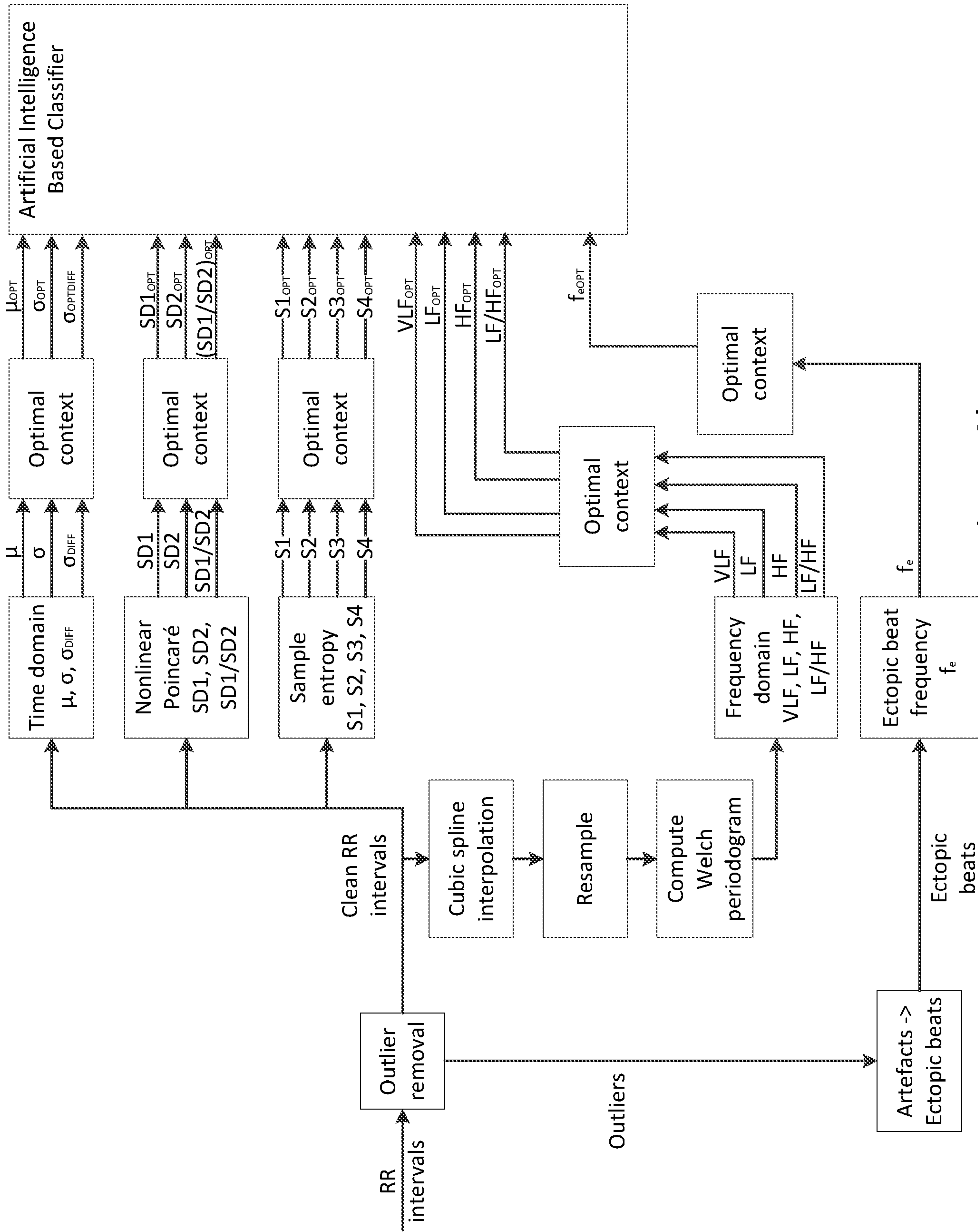


Figure 2b

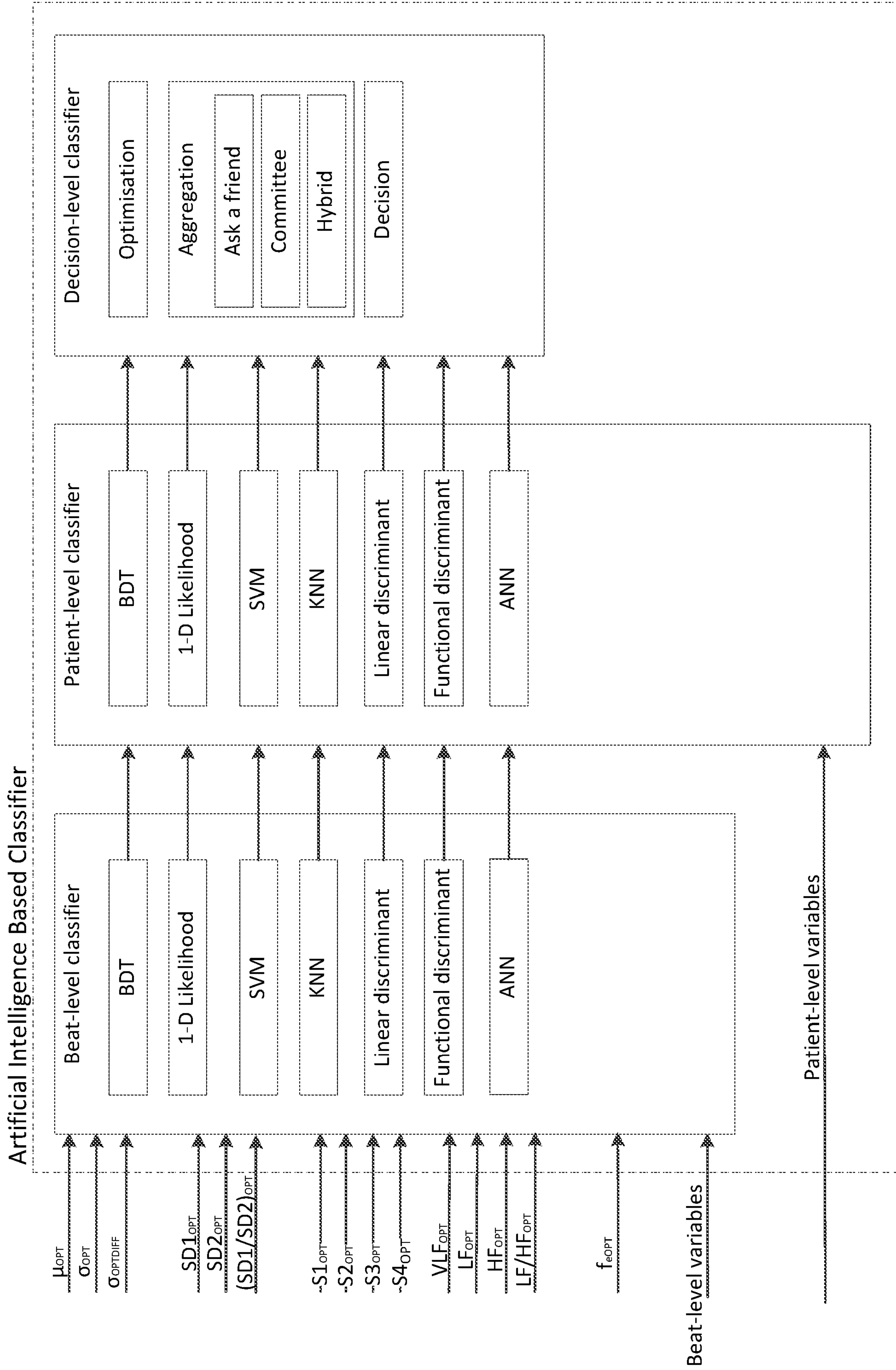


Figure 2c



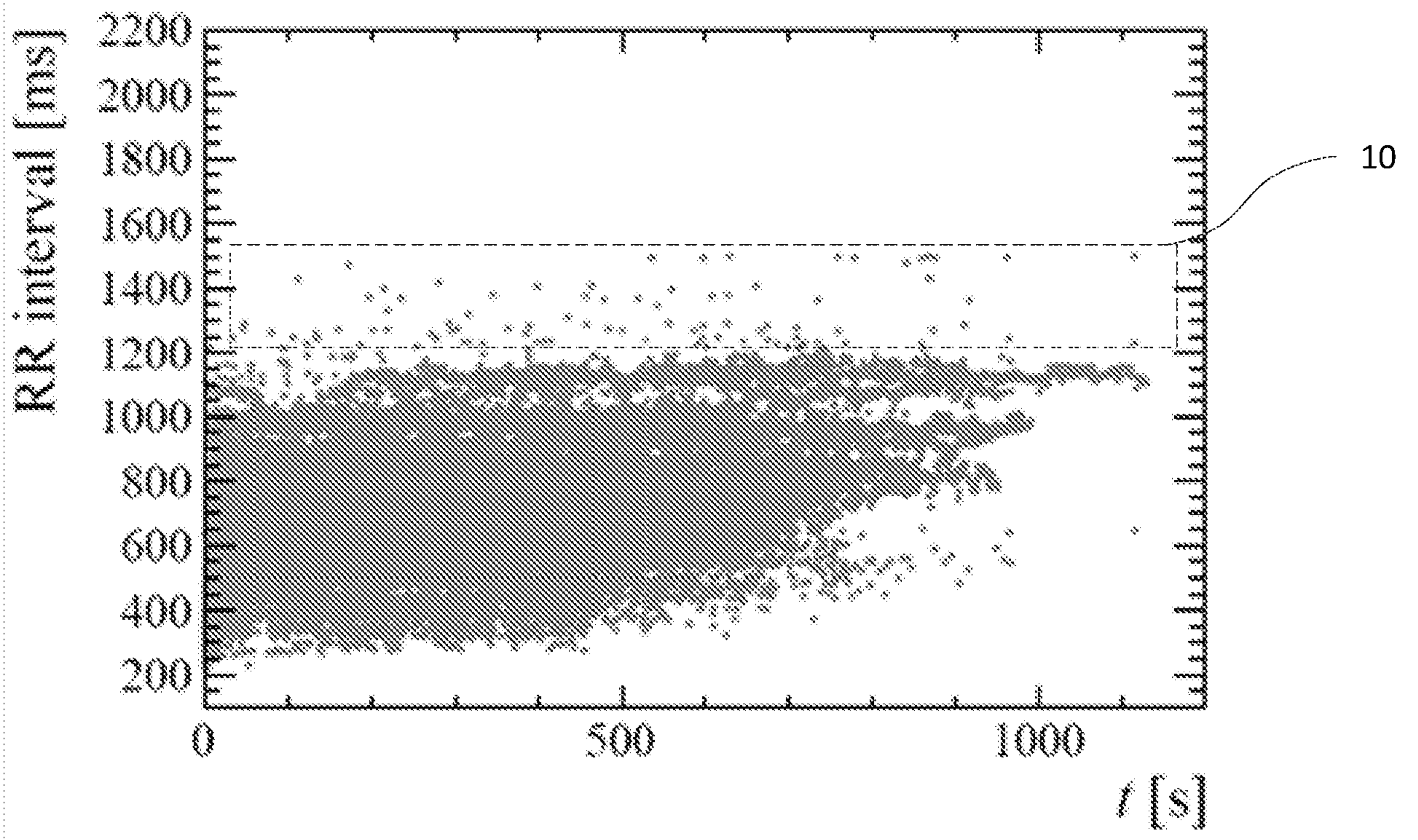


Figure 3a

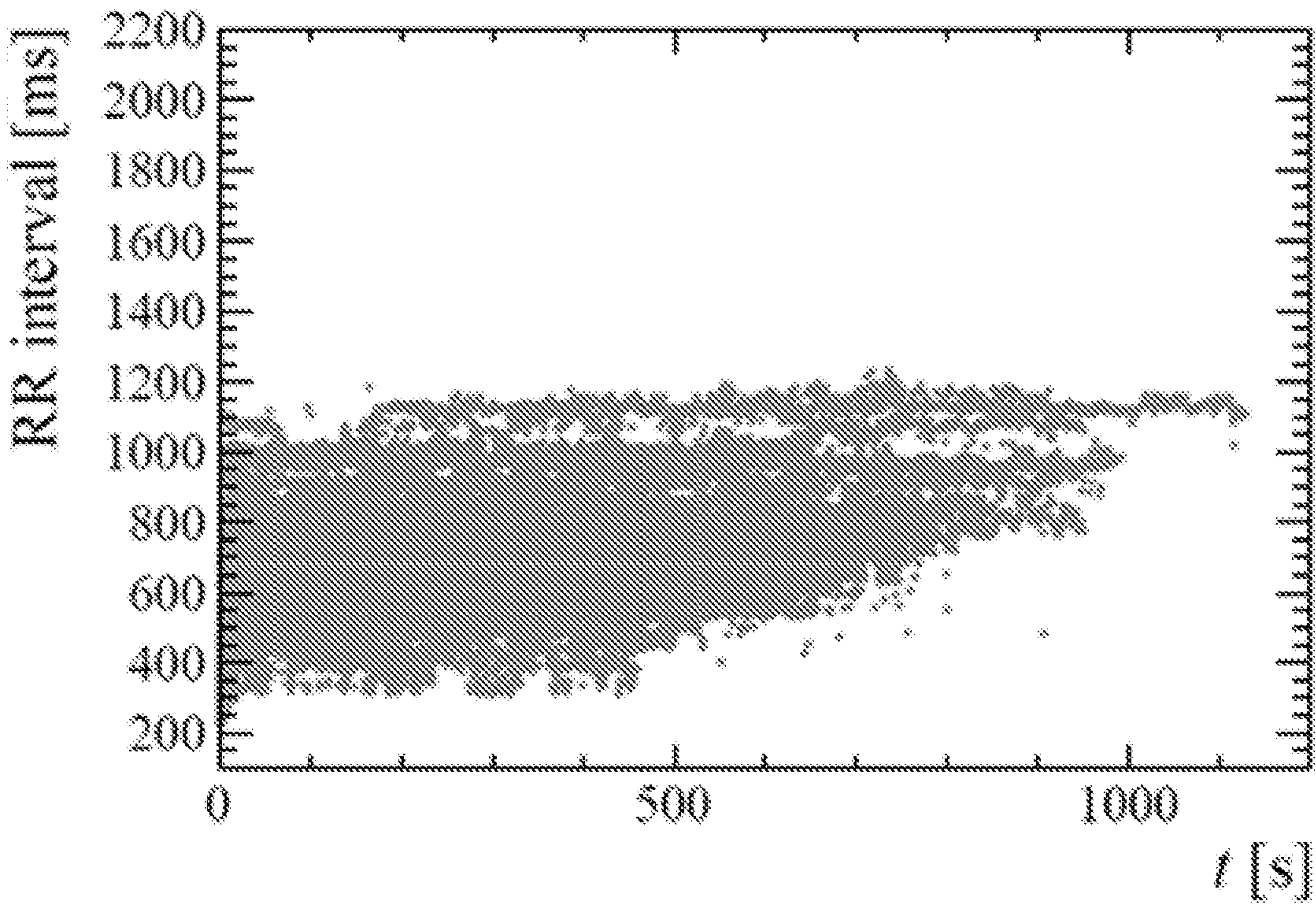


Figure 3b



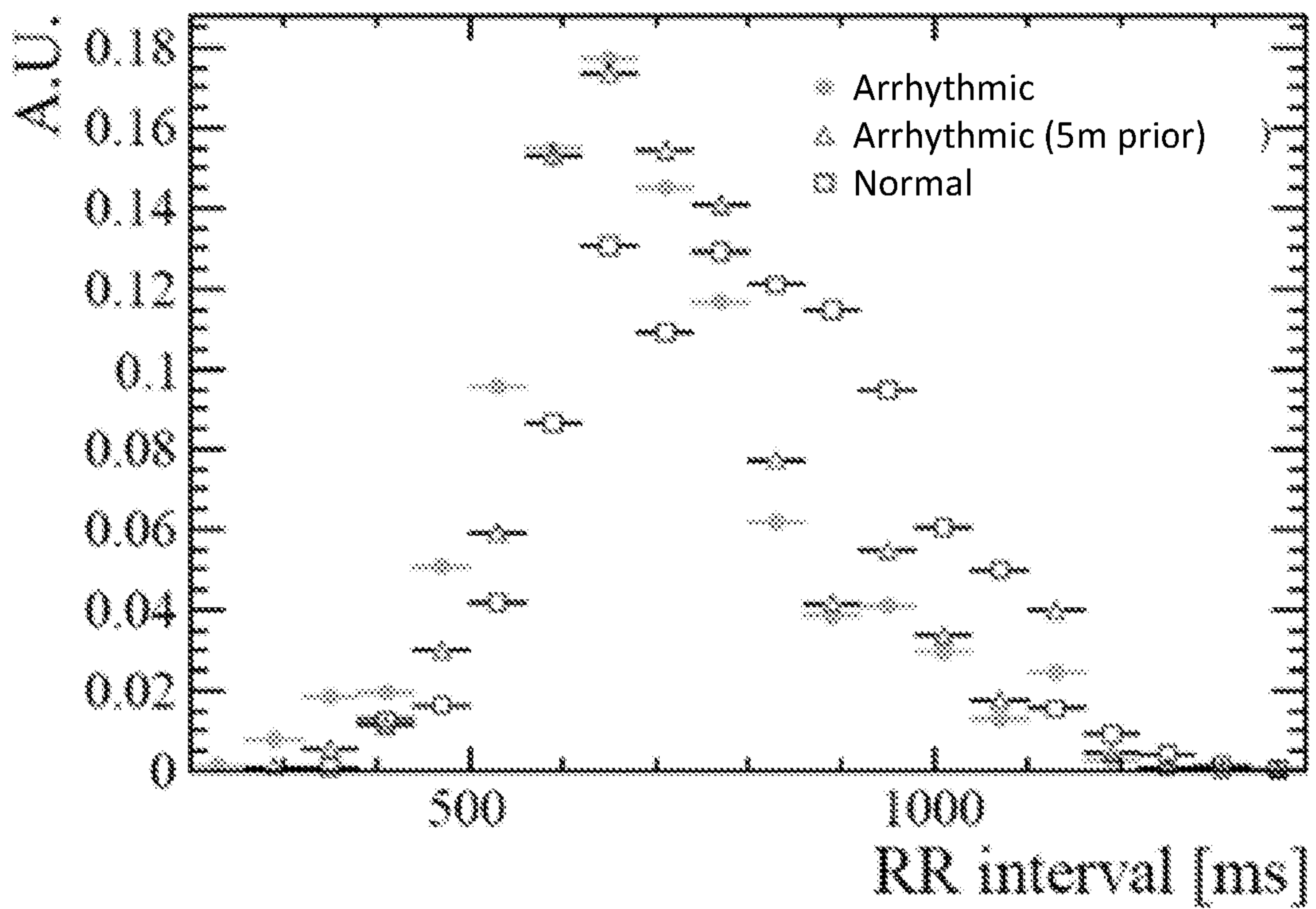


Figure 4a

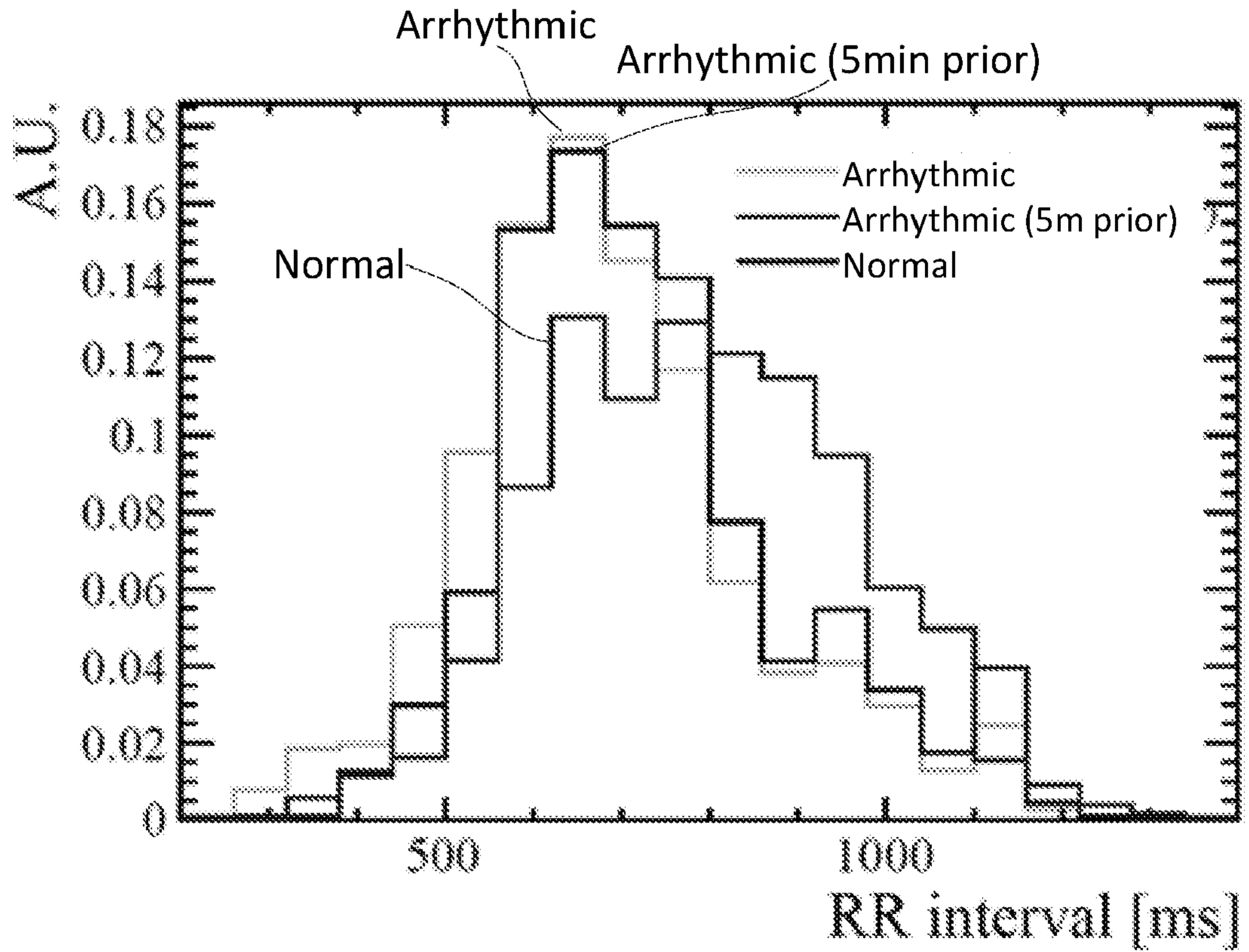


Figure 4b

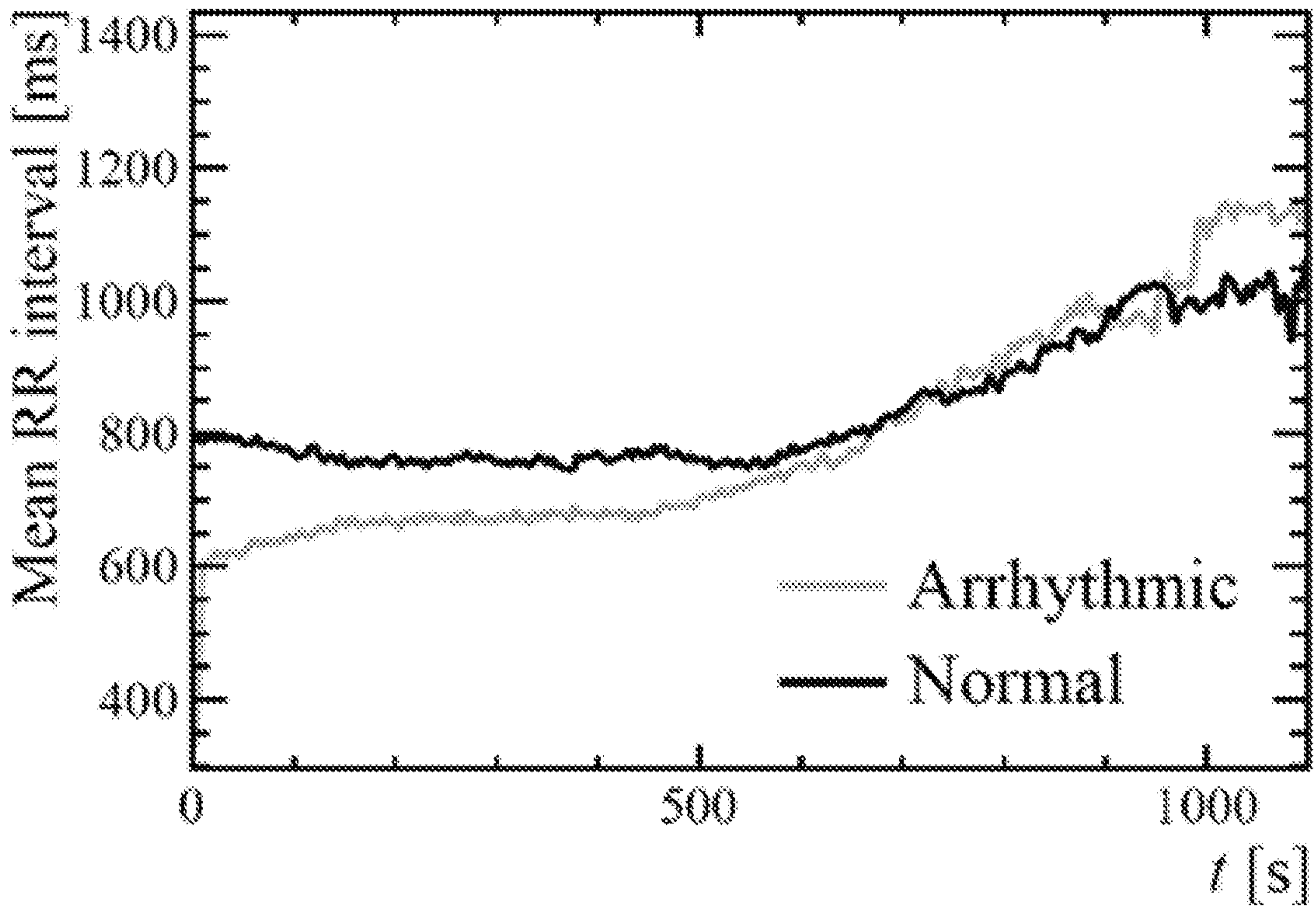


Figure 5

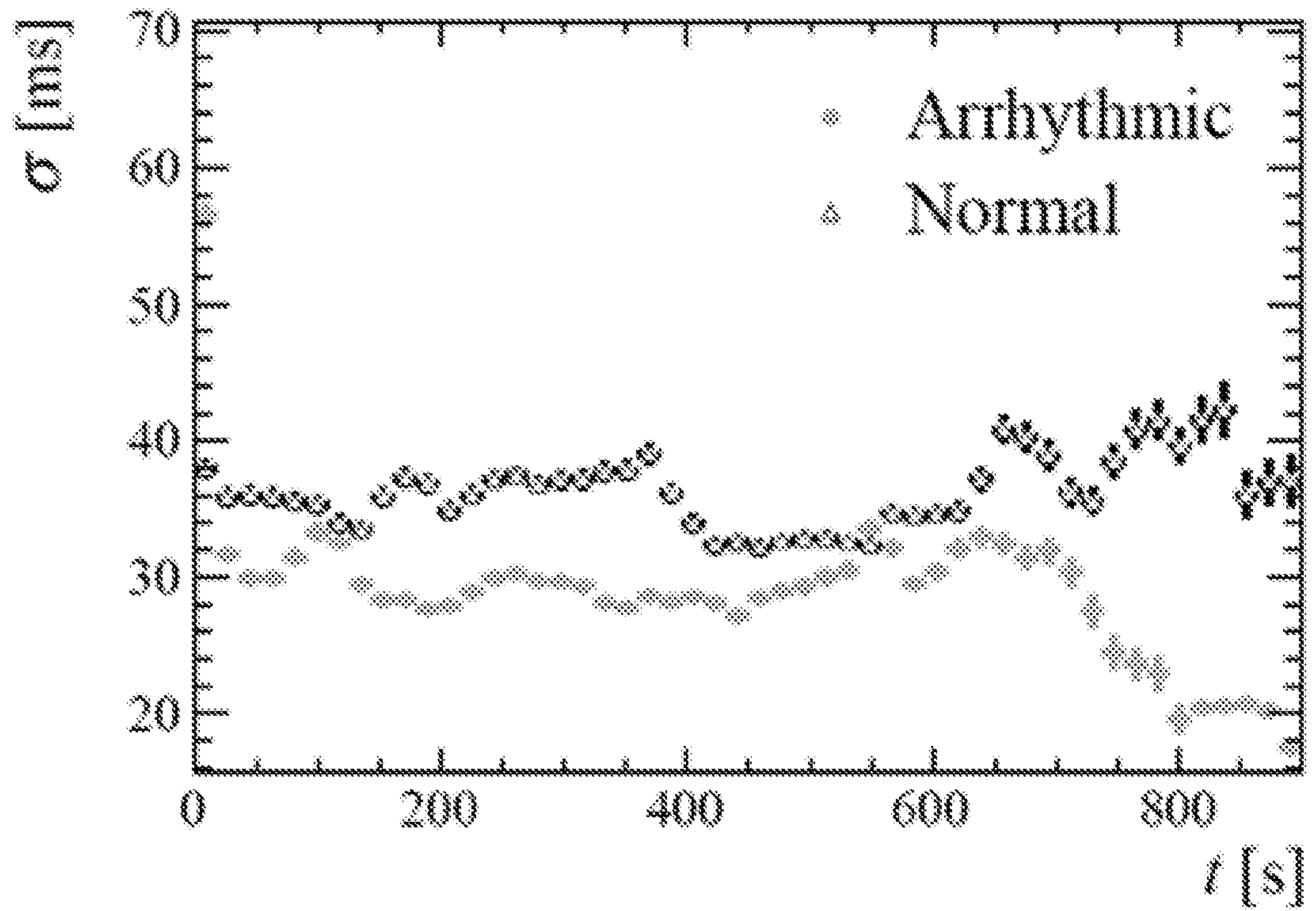


Figure 6

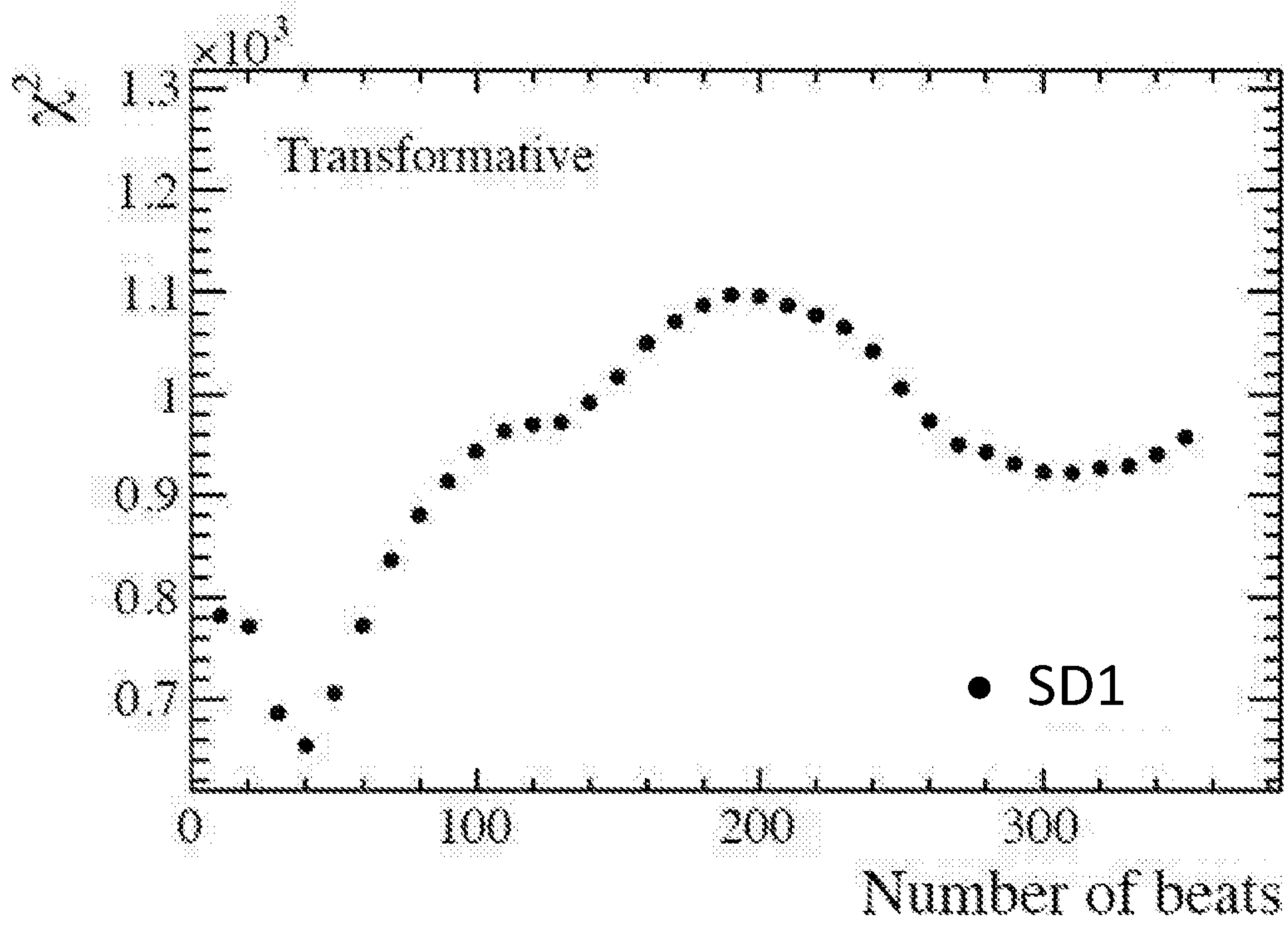


Figure 7a

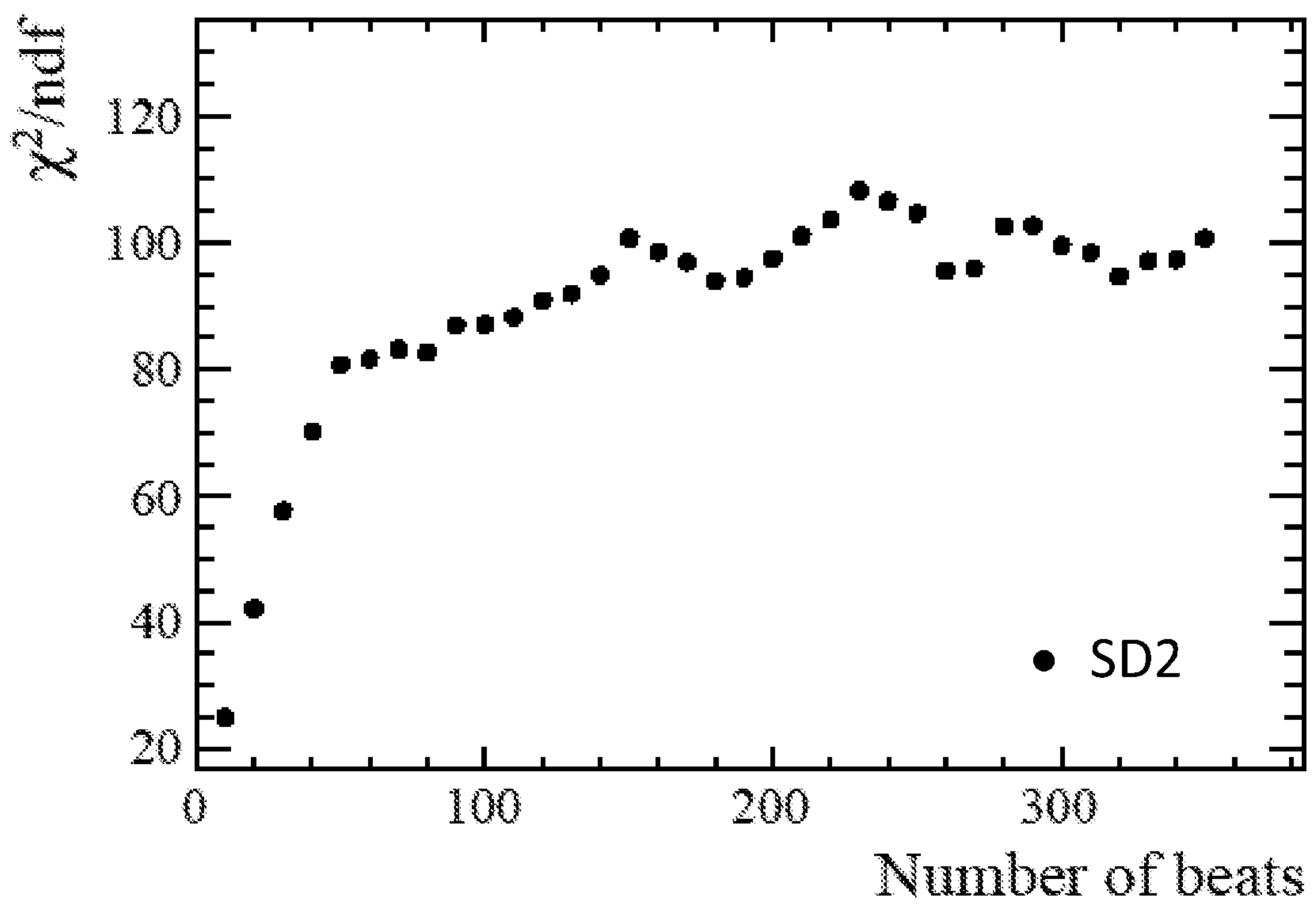


Figure 7b



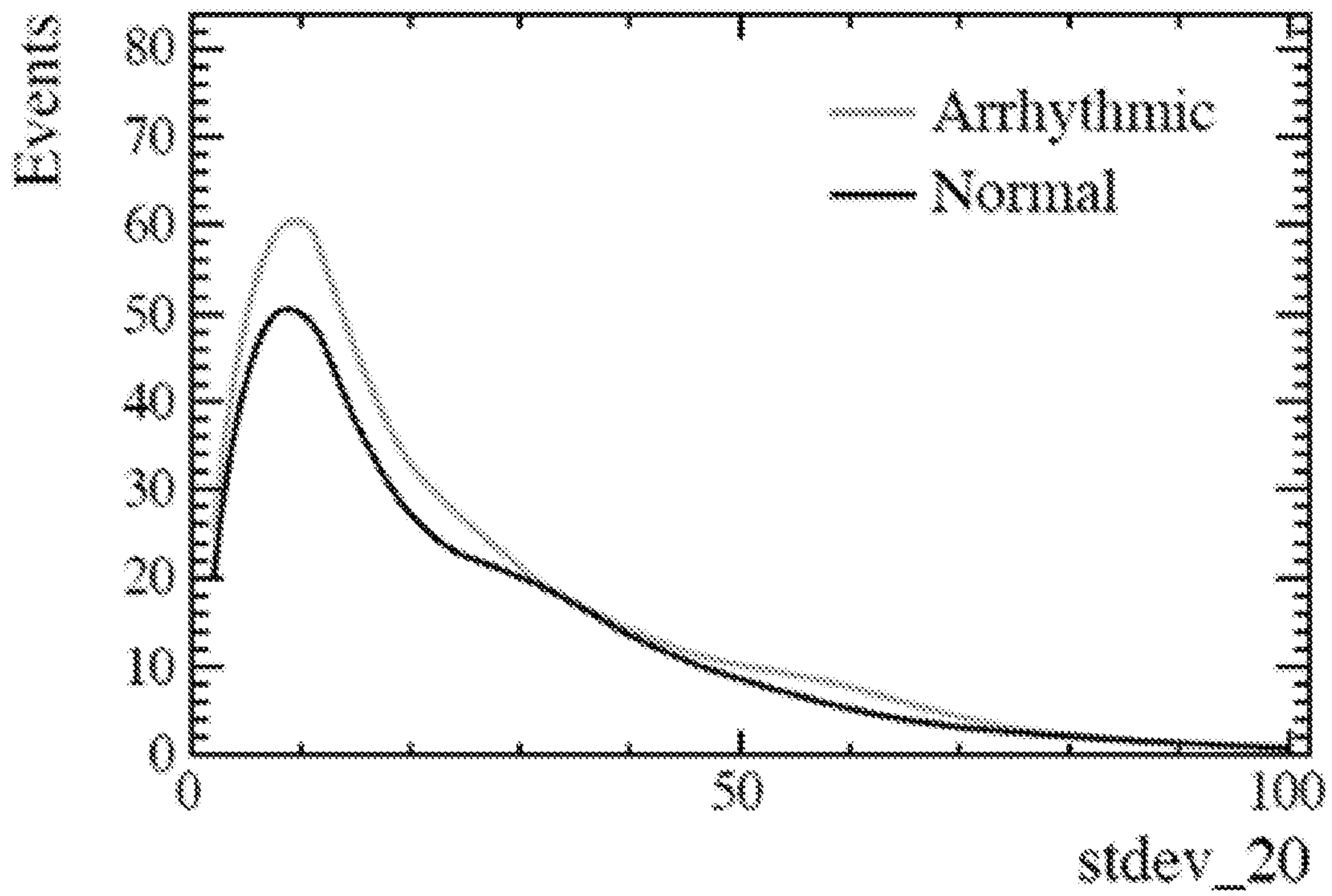


Figure 8a

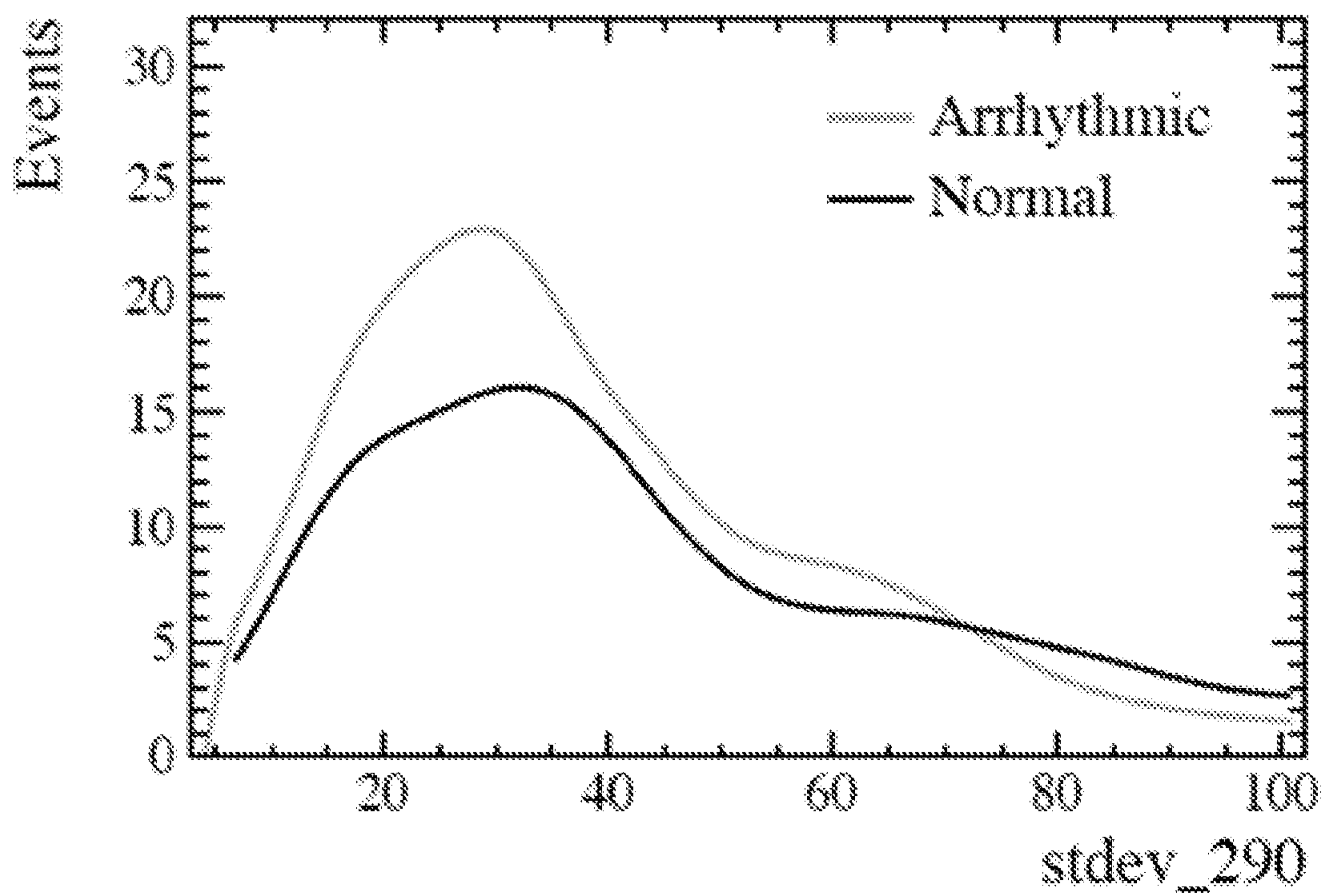


Figure 8b

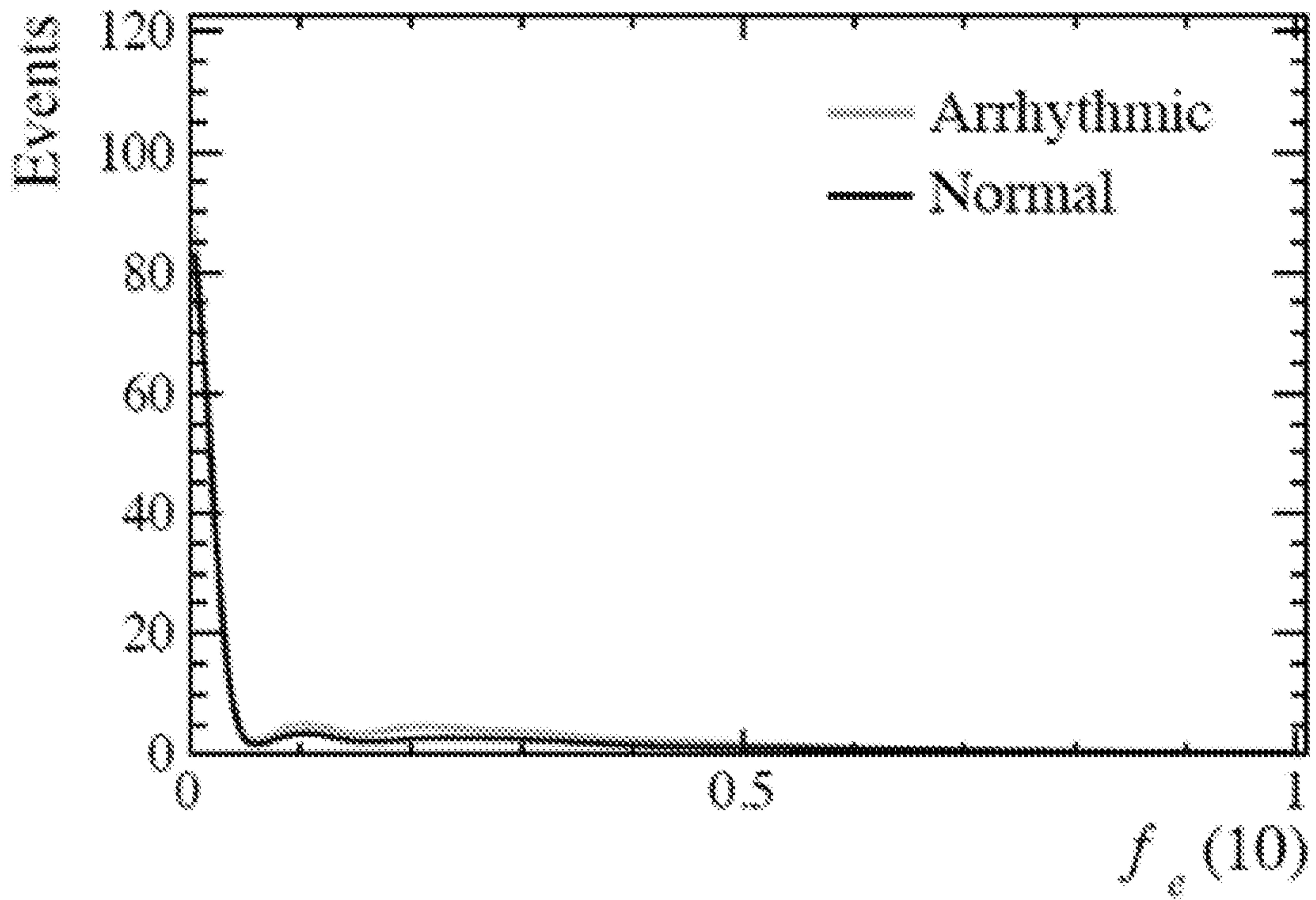


Figure 9a

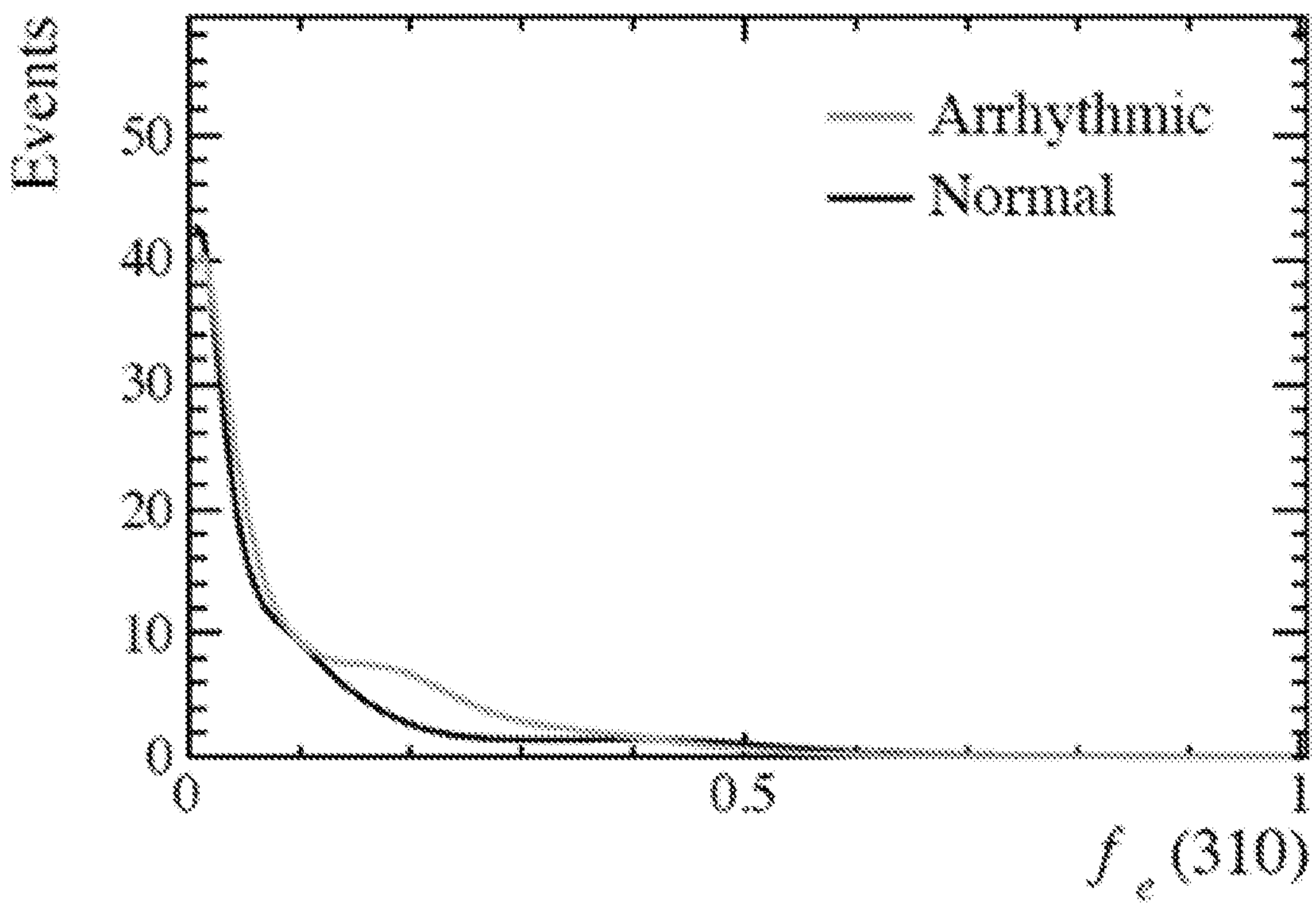


Figure 9b

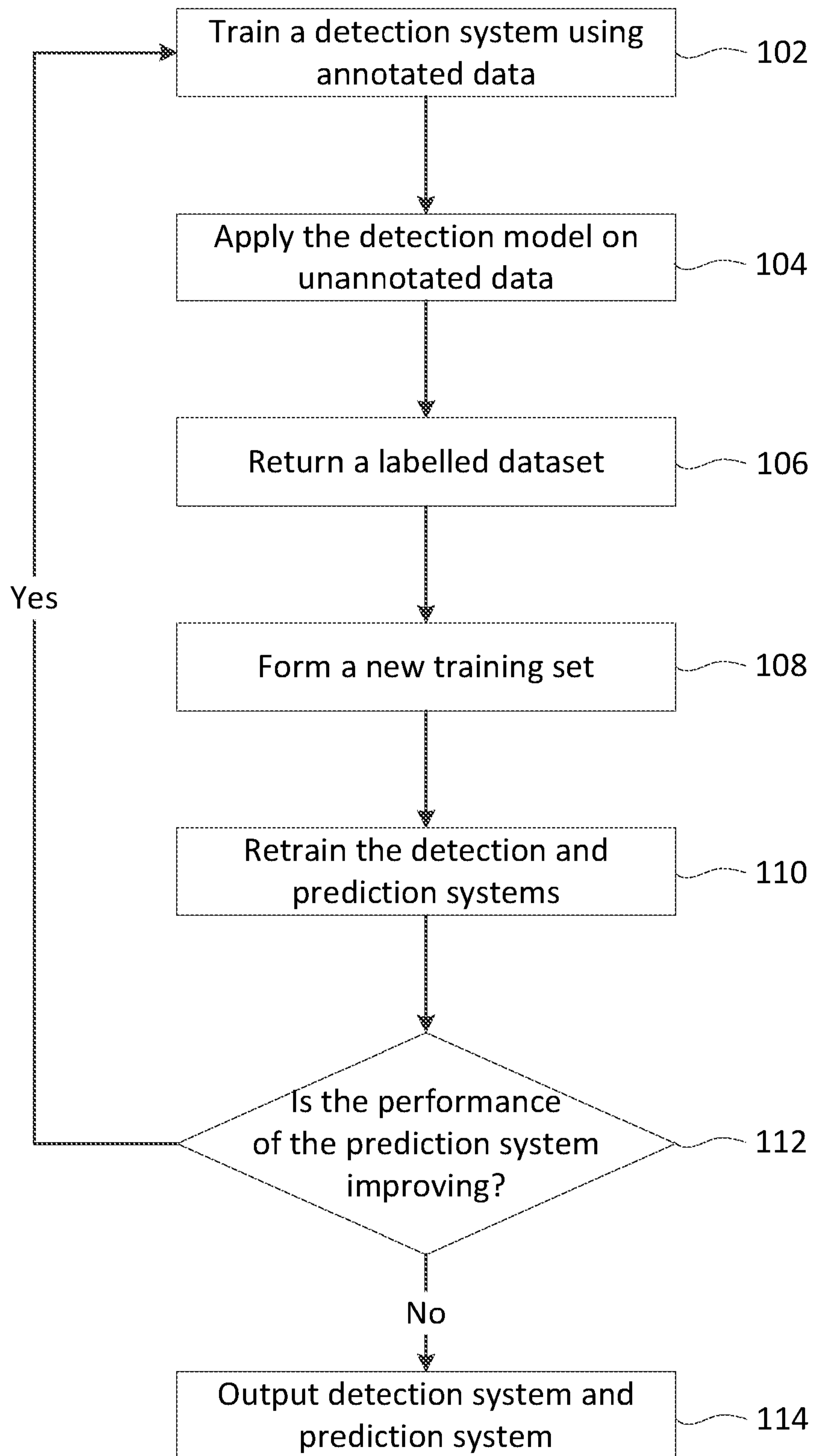


Figure 10



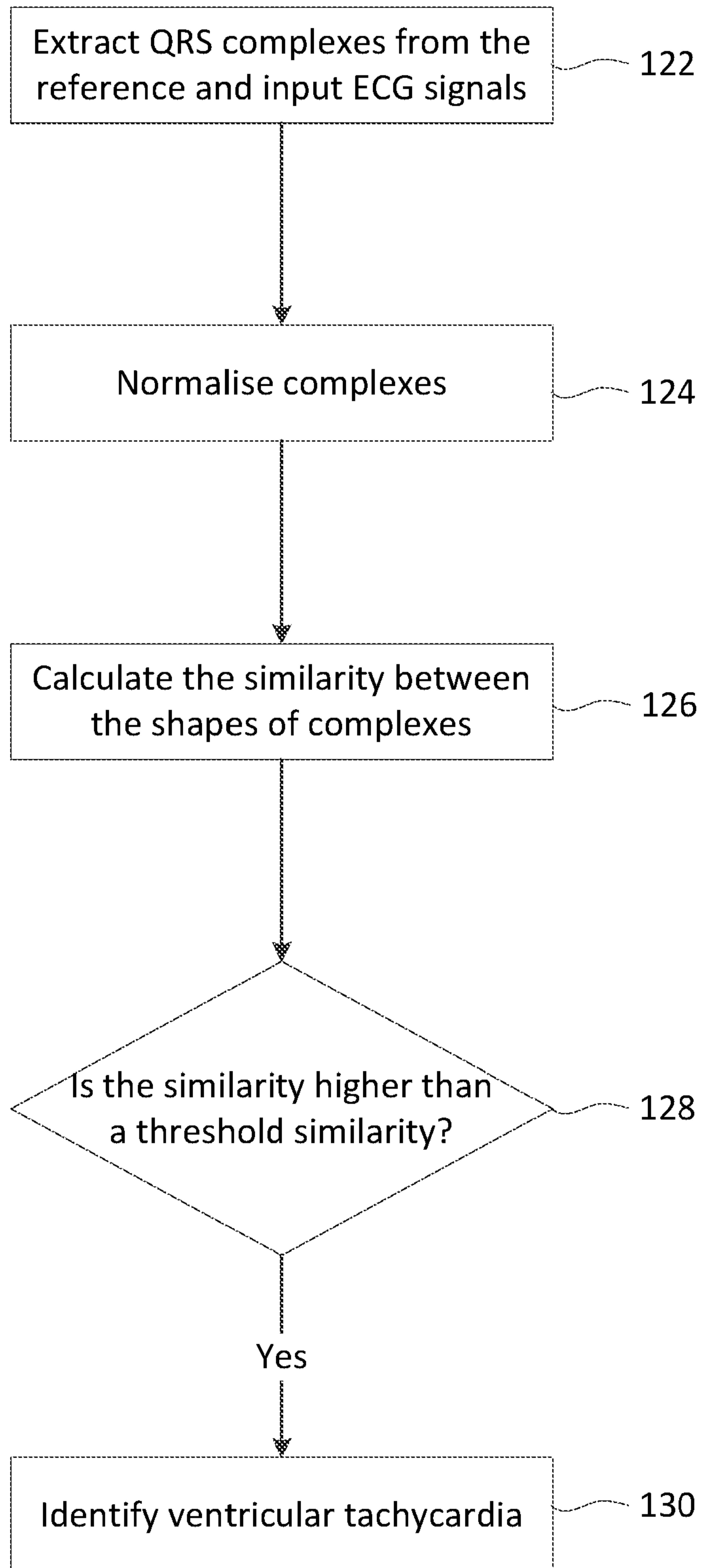


Figure 11

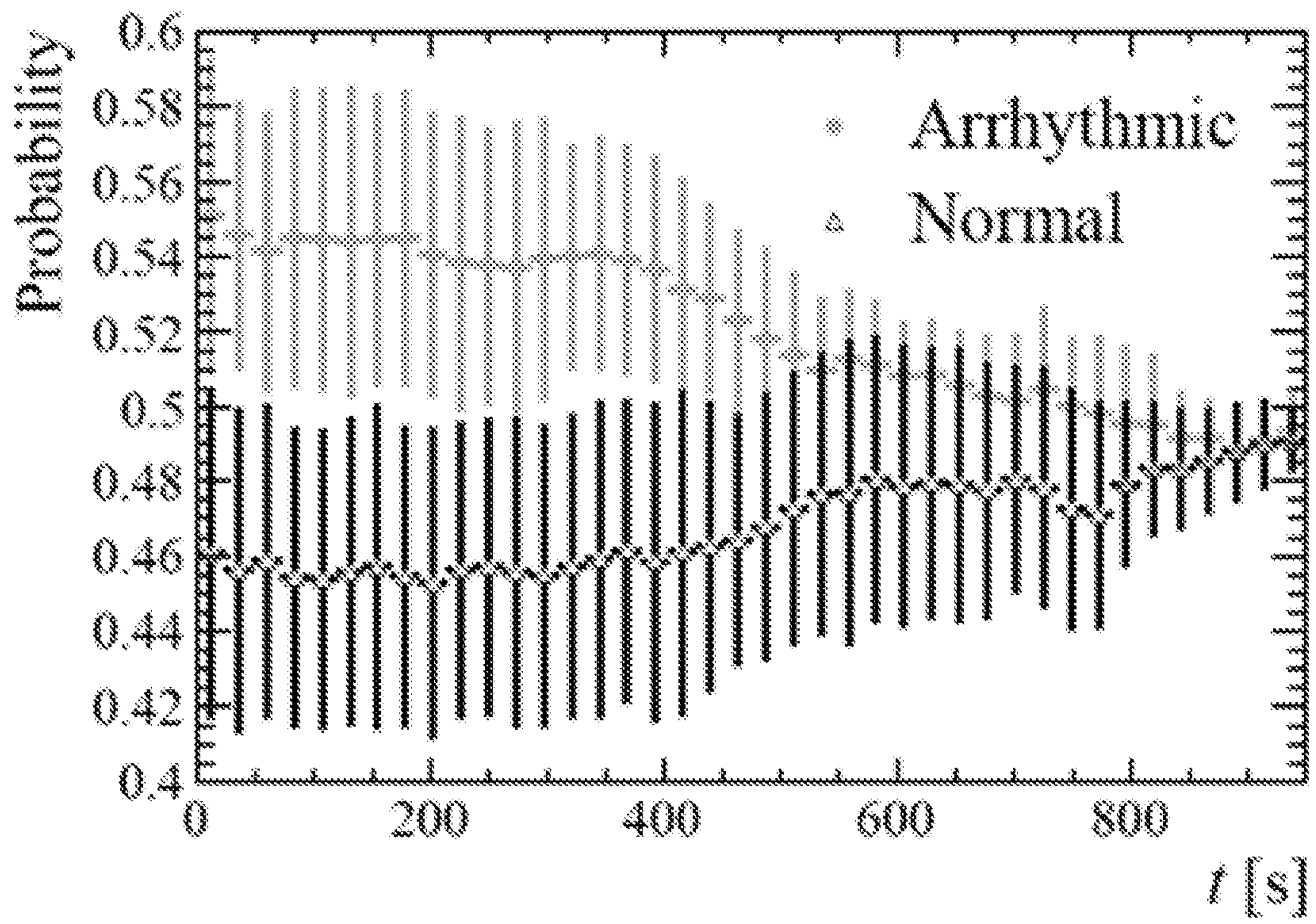


Figure 12

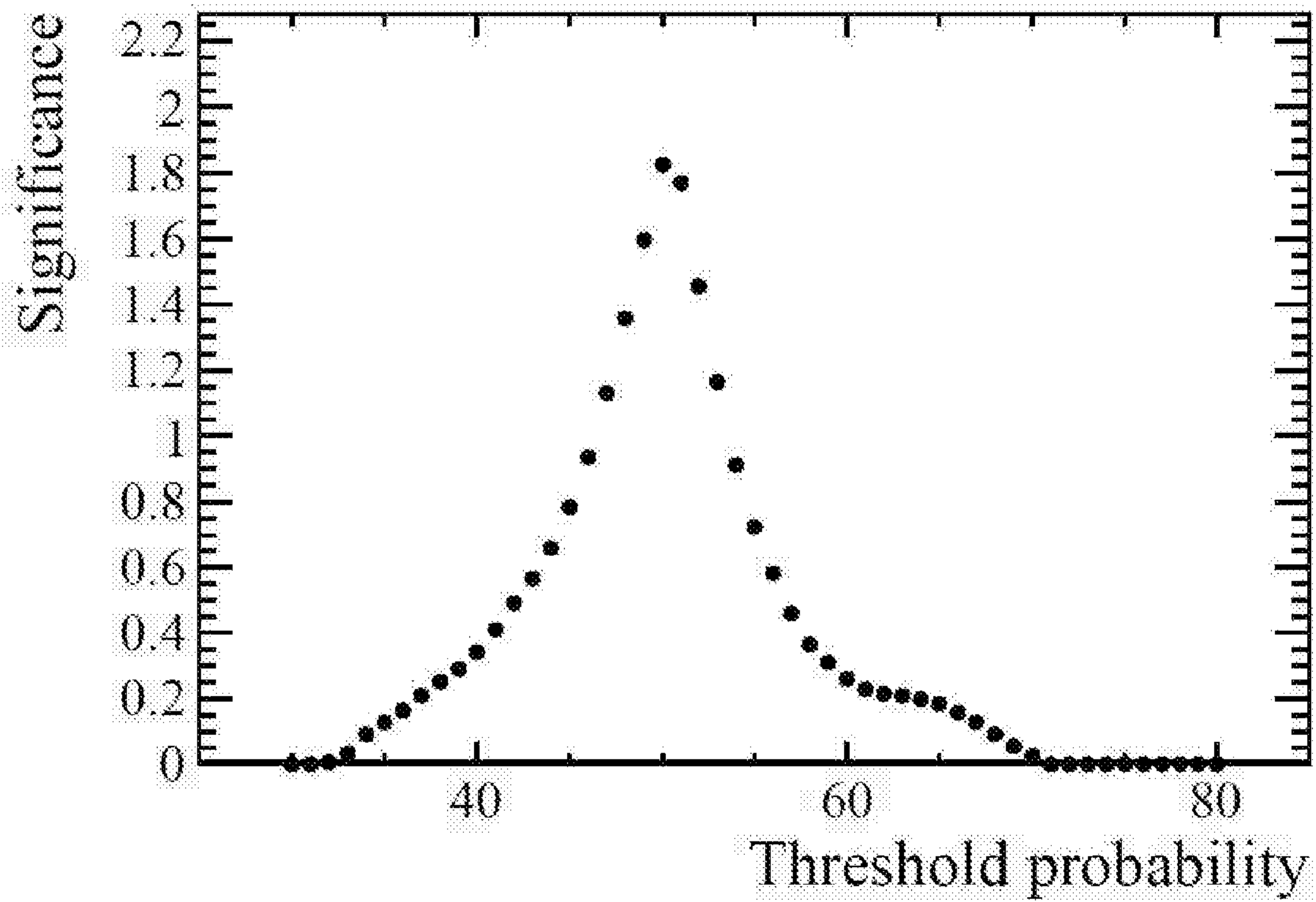


Figure 13

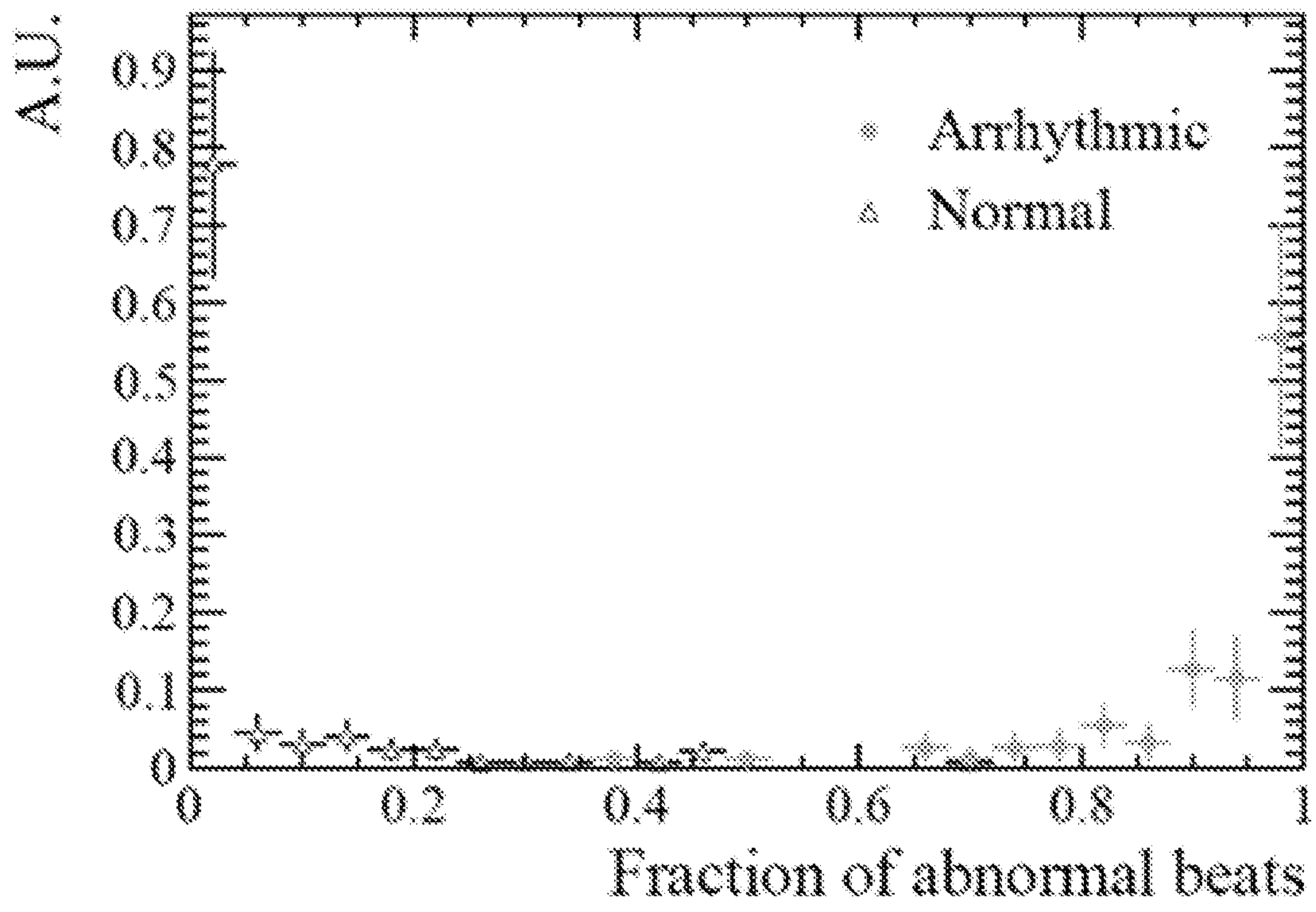


Figure 14



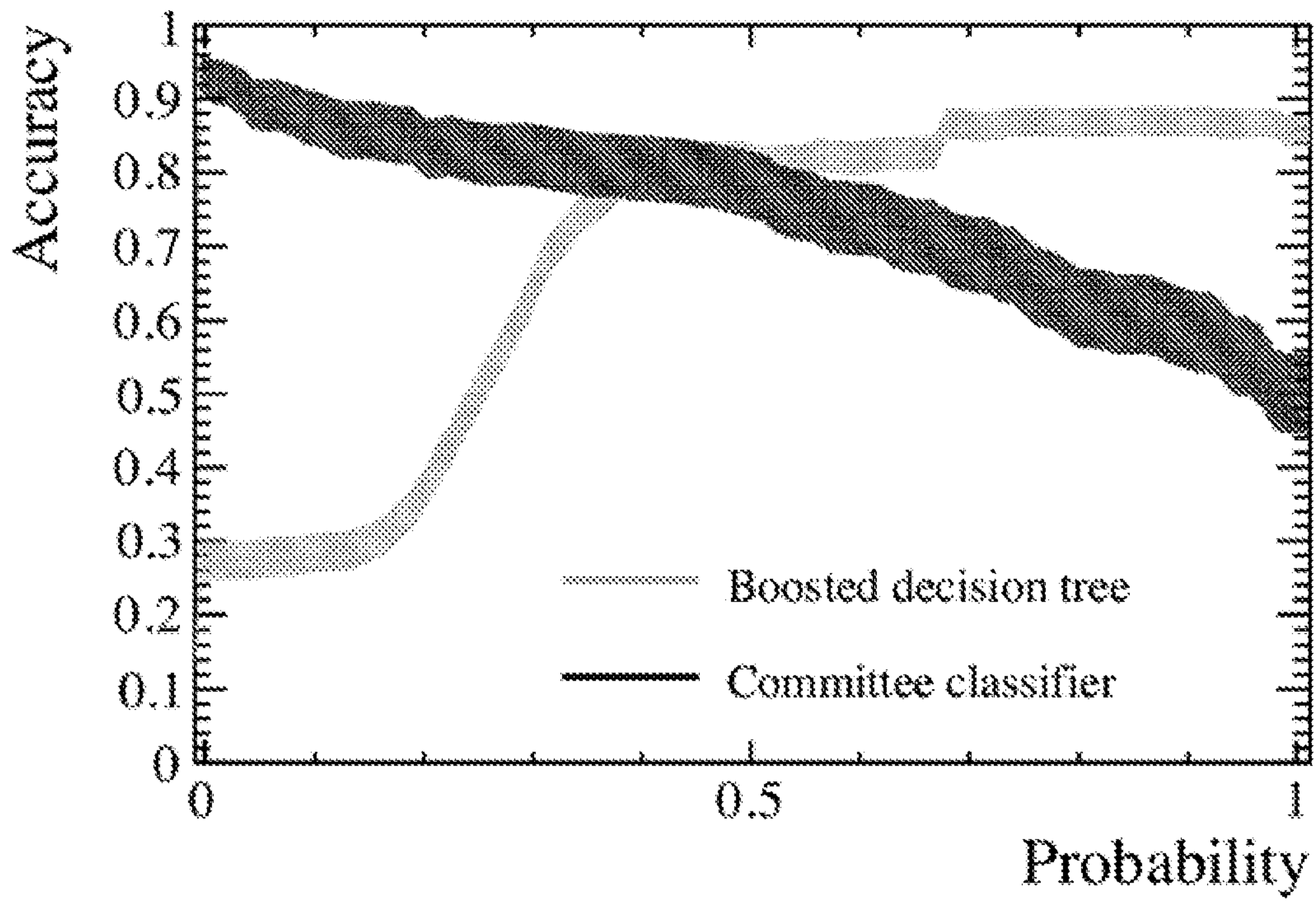


Figure 15a

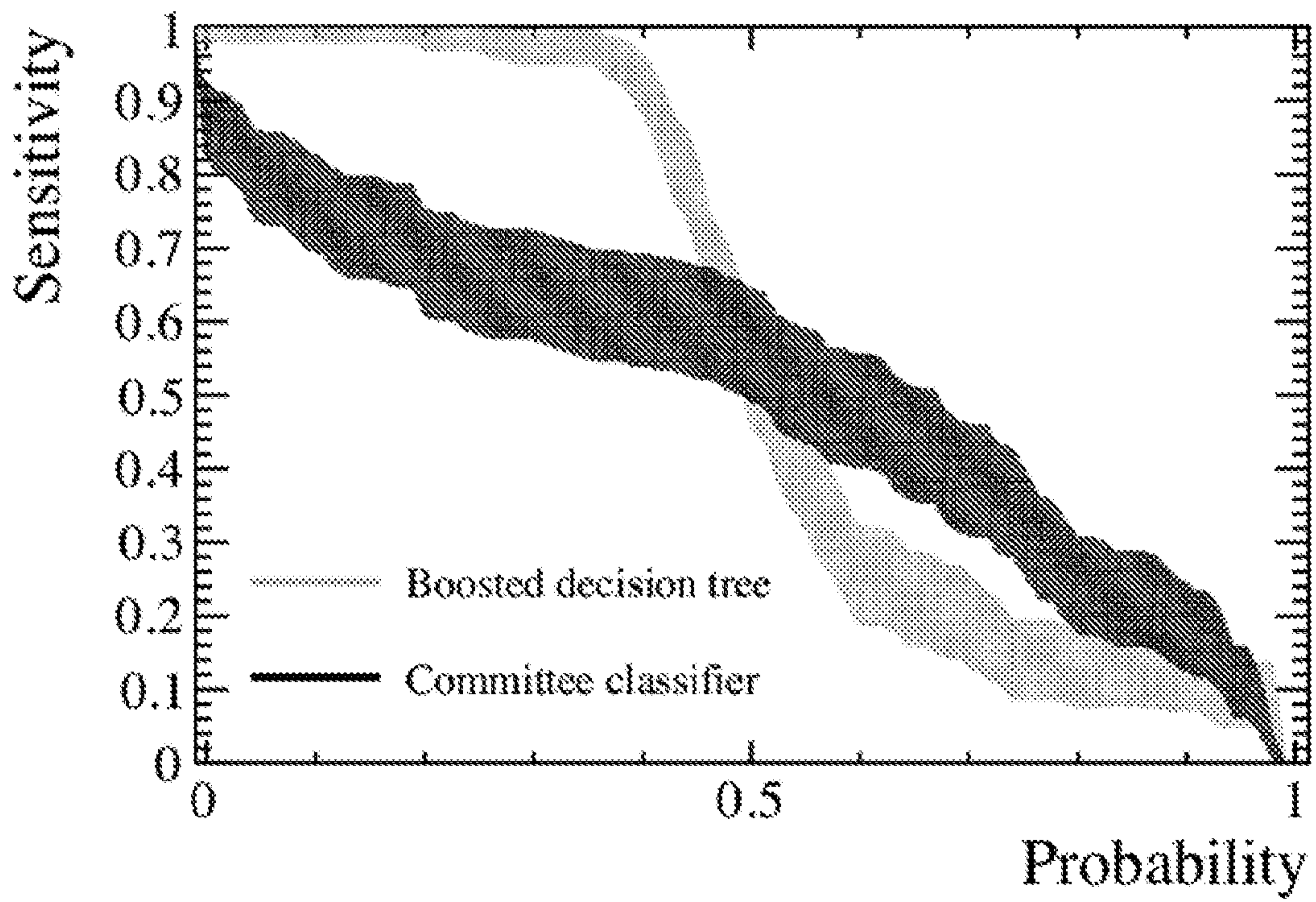


Figure 15b

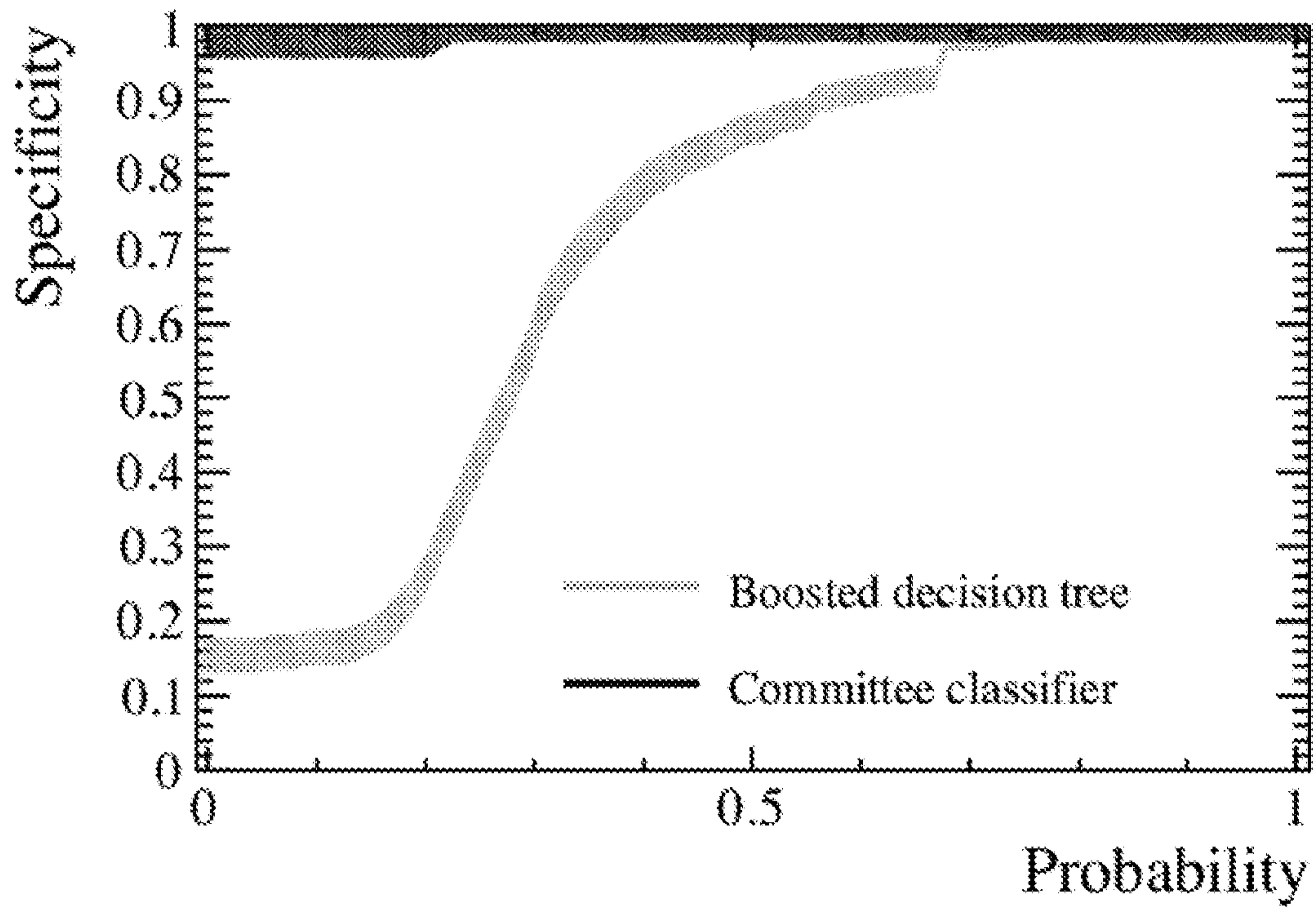


Figure 15c

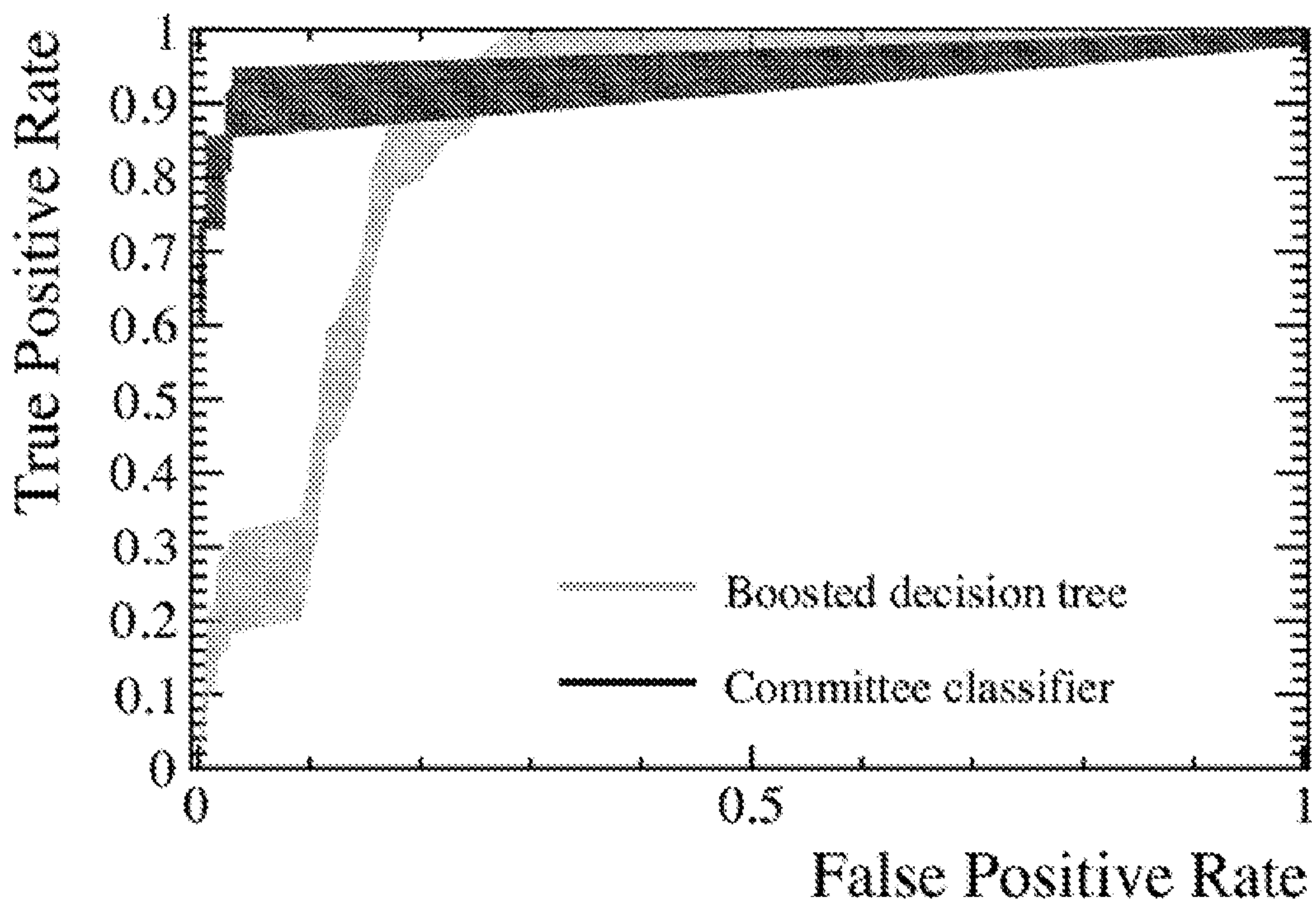


Figure 15d



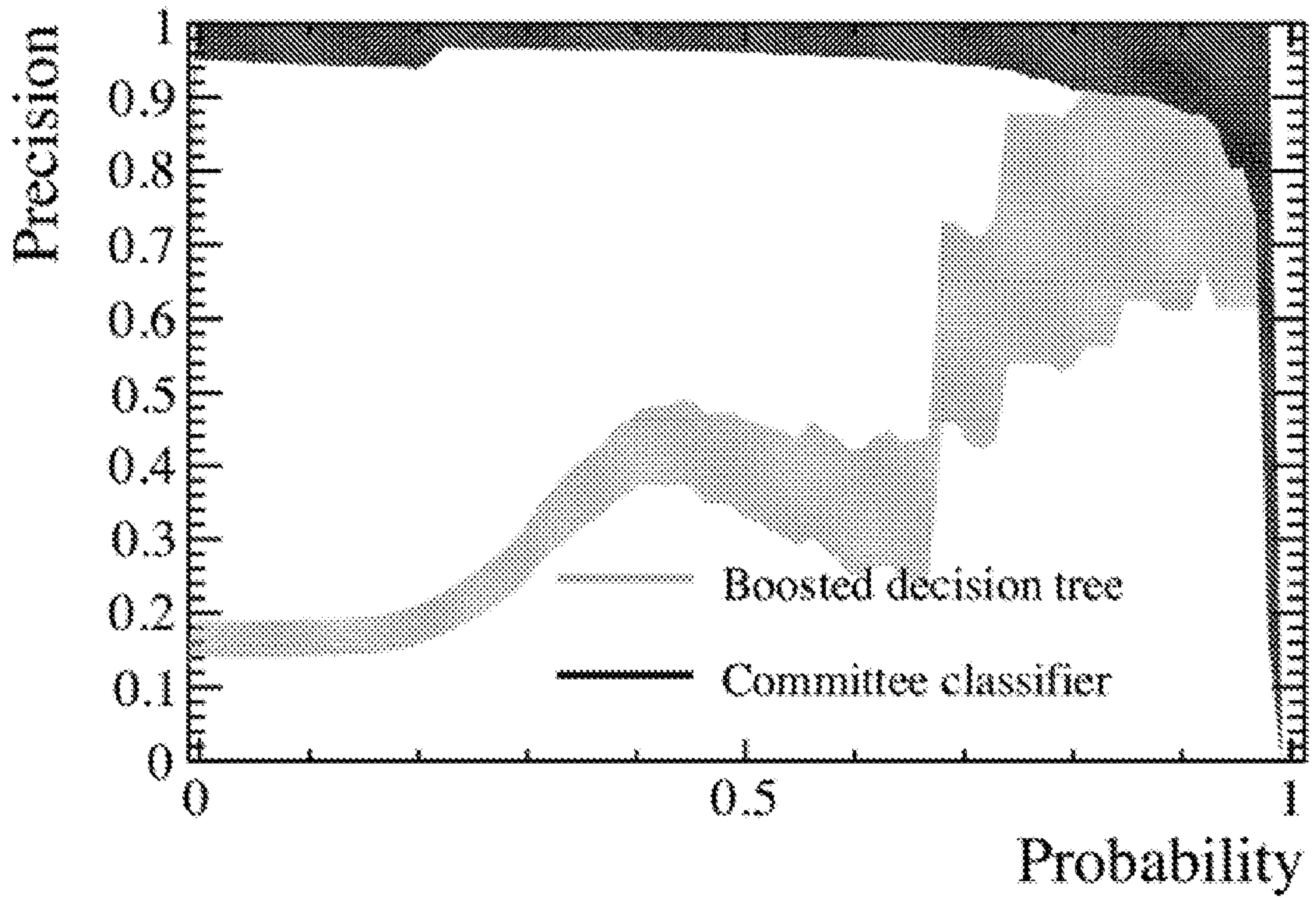


Figure 15e

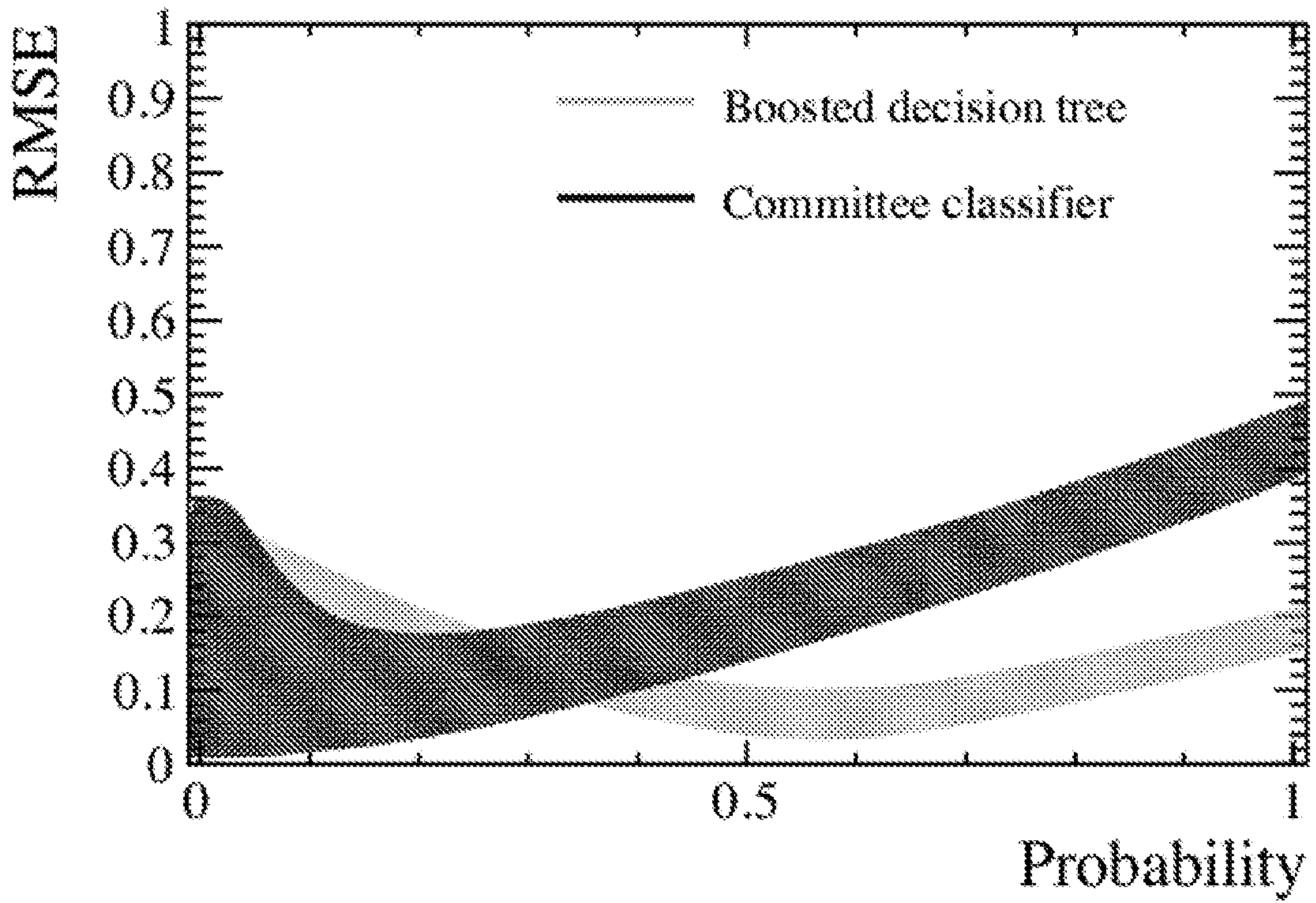


Figure 15f



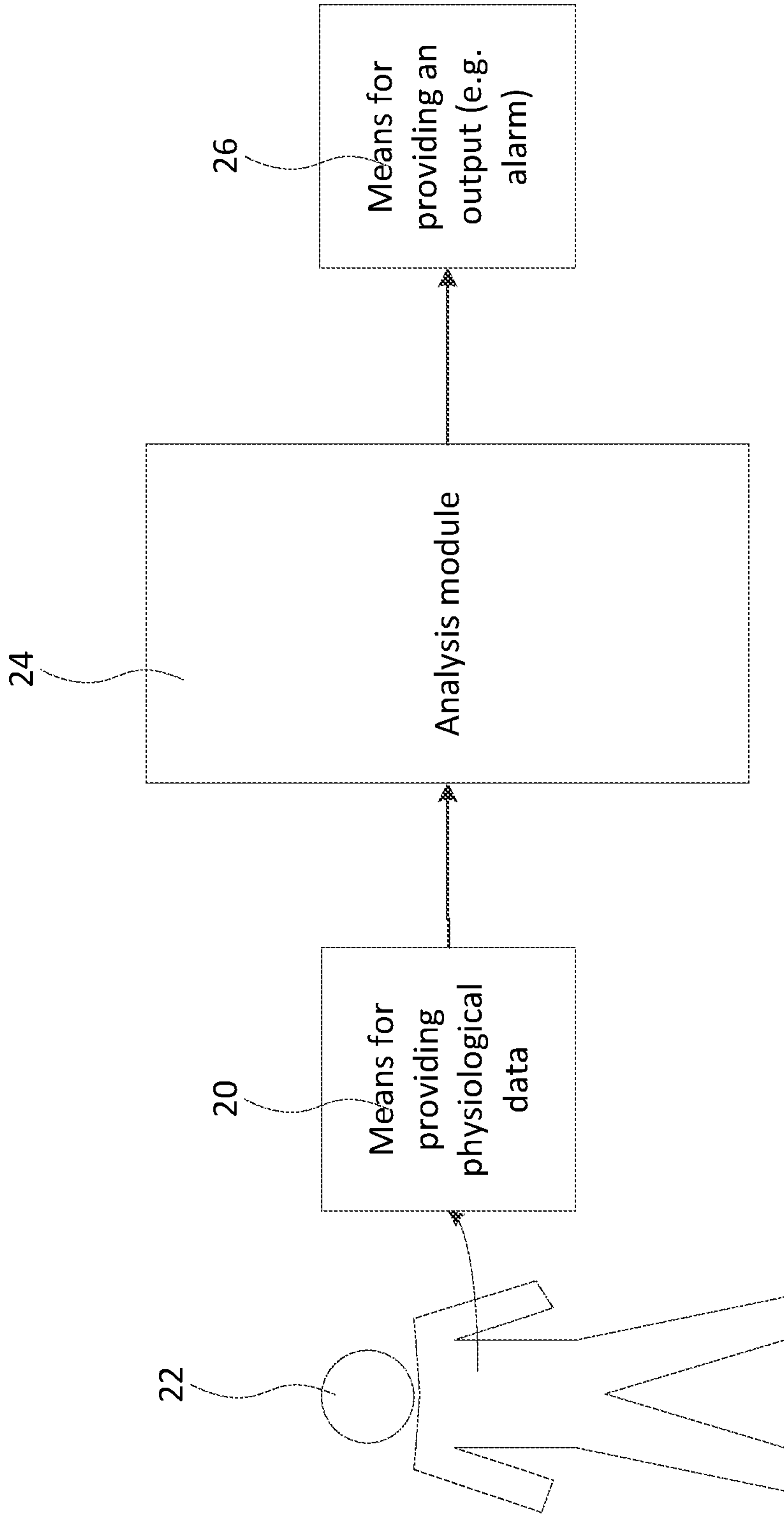


Figure 16

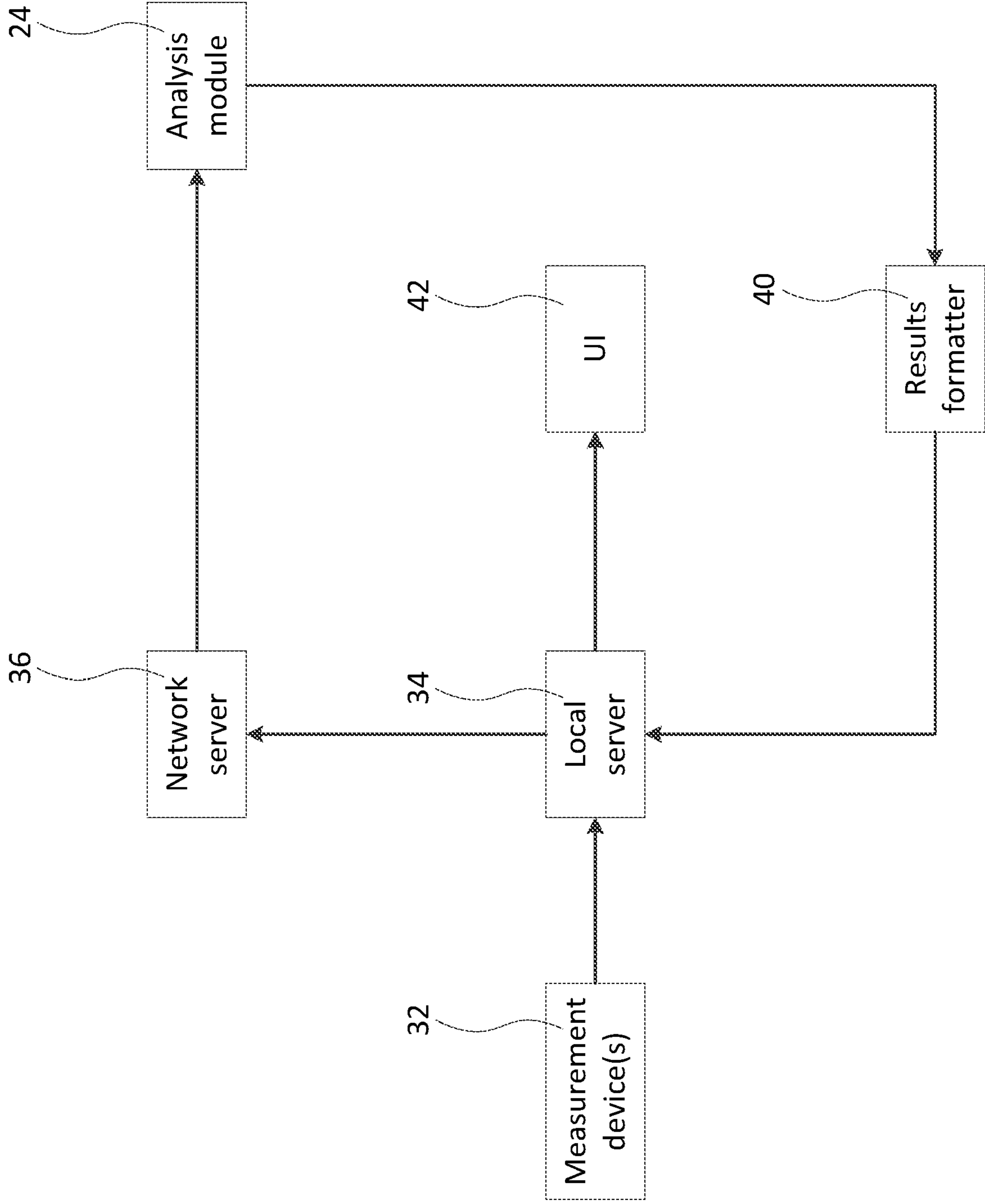


Figure 17

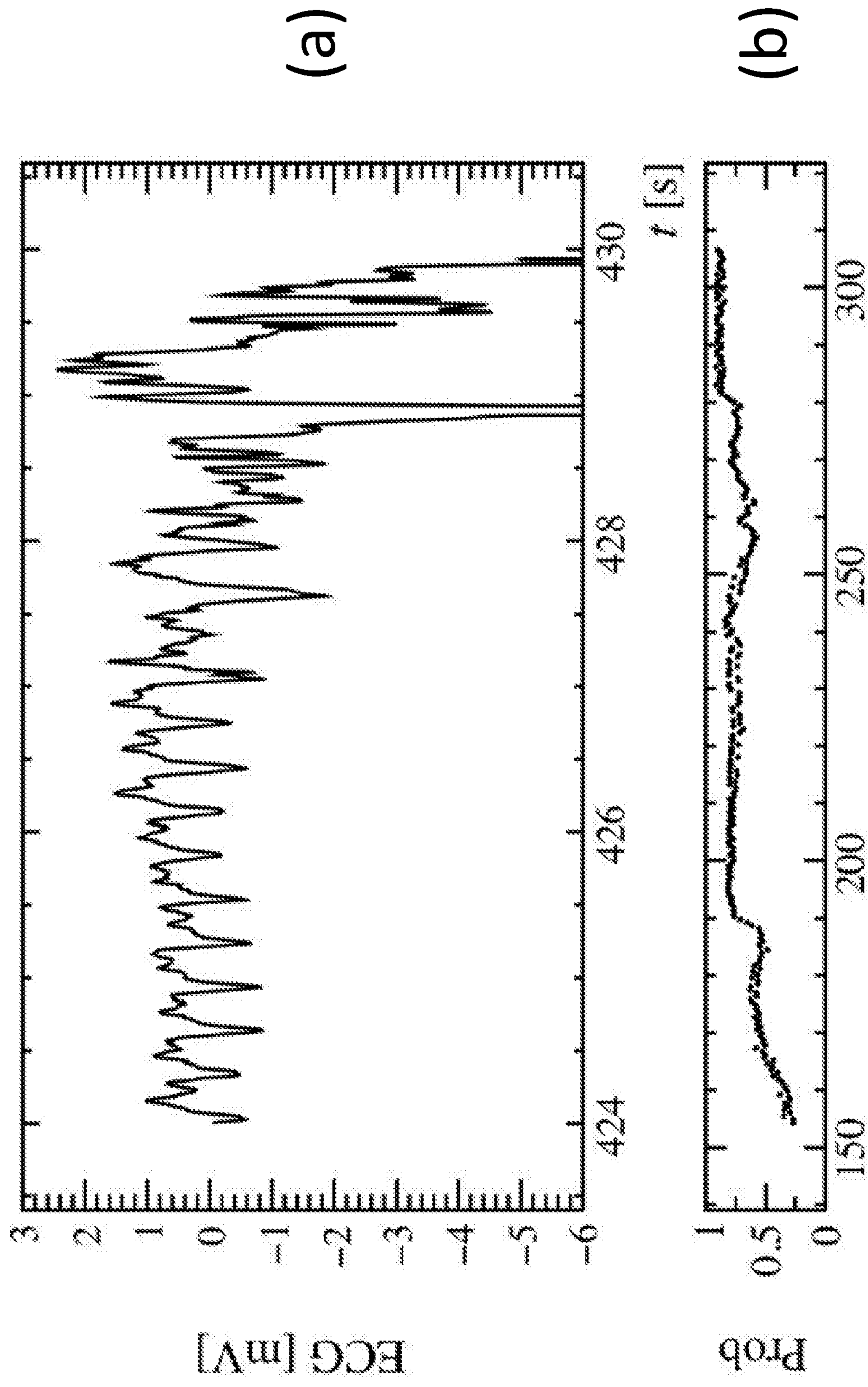


Figure 18





The following terms are registered trade marks and should be read as such wherever they occur in this document:

Fitbit

## Analysis of cardiac data

The present invention relates to the analysis of cardiac data. In particular, the invention relates to a system for and method of determining the probability of a cardiac event occurring. This  
5 may enable timely preventative action to be taken, or may enable the determination of periods during which increased monitoring of a patient would be beneficial.

At present, many patients are monitored in hospital wards and other settings, where measurements of the electrical activity of the heart, such as electrocardiograms, are taken on a  
10 regular basis, sometimes continuously. These measurements enable characteristics of cardiac activity to be extracted. However, only minimal processing is applied to the data, resulting in healthcare providers being confronted with large amounts of noisy raw data. While current technology allows for algorithmic analysis and description of the patient's cardiac activity at the time of analysis (and potentially allows for determination of a long-term risk), it does not provide  
15 predictive analysis of short to medium term potential future cardiac activity. The present disclosure aims to make the measured data more useful and actionable through sophisticated data analysis and the use of artificial intelligence, so that cardiac events (and more generally periods of increased risk, where additional attention should be given to a patient) may be predicted.

20 Aspects and embodiments of the present invention are set out in the appended claims. These and other aspects and embodiments of the invention are also described.

Described herein is a method of predicting cardiac events using heart rate data relating to a  
25 patient, comprising [the steps of]: evaluating a property of multiple heartbeats within said heart rate data; determining a value associated with the number of said multiple heartbeats that exceed an abnormality threshold set for said property; and comparing said value against a predetermined value [for a given time window], thereby to indicate a probability of said patient experiencing a cardiac event; wherein the abnormality threshold is determined based on a  
30 dataset of a plurality of heart rate data obtained from multiple sources.

Also described herein is a method of predicting cardiac events using physiological data relating to a patient, comprising: inputting physiological data relating to the patient; evaluating a property of multiple heartbeats within the physiological data; determining a value associated  
35 with the number of the multiple heartbeats that exceed an abnormality threshold set for said property; comparing the value against a predetermined value, thereby to indicate a probability of the patient experiencing a cardiac event; and providing an output based on the comparison; wherein the abnormality threshold is determined based on a dataset of a plurality of

physiological data obtained from multiple sources.

5 Providing a method of predicting events that are likely to lead to medical conditions enables timely preventative action to be taken. In particular, the method described herein may be used to predict the onset of arrhythmias that are likely to lead to cardiac events (e.g. Sudden Cardiac Arrest). As described herein, this may be achieved by using physiological data to produce a single probabilistic assessment of the likelihood of a patient experiencing a cardiac event in a subsequent time period. The applicable time scales are considered to be from seconds to hours in advance of the episode. The present invention may help to reduce the workload for healthcare providers by reducing the volume of data they are confronted with. It may also improve outcomes for patients and lead to fewer complications.

15 Also described herein is a method of analysing cardiac data relating to a patient, comprising: providing cardiac data relating to the patient; determining one or more properties of the data, wherein each property is determined over a particular context length, the context length being selected based on the property; comparing the or each property against a respective predetermined threshold value, thereby to indicate a probability of the patient experiencing a cardiac event; and providing an output based on the comparison.

20 According to at least one aspect of the present disclosure, there is described a method of analysing cardiac data relating to a patient, comprising providing cardiac data relating to the patient; determining a property of the data, wherein the property is determined over a particular context length, the context length being selected based on the property; comparing one or more features of the property against a predetermined threshold value, thereby to indicate a probability of the patient experiencing a cardiac event; and providing an output based on the comparison.

30 A patient, as specified here, may refer to a patient who is currently receiving medical care, for example in a hospital, but could equally relate to, for example, a person not currently receiving care, where cardiac data is obtainable (for example a person with a defibrillator, who is considered healthy).

35 The method preferably further comprises modelling the property using a function; and comparing one or more descriptors of the function against a predetermined descriptor threshold value. The descriptors may be compared in place of, or in addition to, the features of the property. Preferably, a continuous function is used. Comparing one or more descriptors of the function may comprise comparing at least one of: a mean; a variance; and a kurtosis.



A plurality of datapoints related to the property may be determined and the distribution of the datapoints modelled using a function.

- 5 Preferably, modelling the property using a function comprises determining a probability density function suitable for modelling the property and/or superposing one or more Gaussian functions, preferably superposing Gaussian functions of equal surface.

10 Preferably, the method further comprises providing contextual data relating to the patient; wherein the threshold value is dependent upon the contextual data. This contextual data enables the comparison to take into account the background of the patient when considering the data.

15 Optionally, the method further comprises comparing a further property against a contextual threshold value, wherein the contextual threshold value is dependent upon contextual data; and providing an output based on both the comparison of the property and the comparison of the further property. Optionally, the comparison of the further comparison is dependent upon the comparison of the property, where this may be used as a secondary comparison that is only performed if the first comparison suggests there is a heightened risk of a  
20 cardiac event. Only performing the secondary comparison under certain circumstances may save computing power. Optionally this may lead to two outputs, where the first comparison may result in a first warning and the second comparison may result in a second warning.

25 Contextual data may comprise: historic data related to the patient, an electronic health record related to the patient, physical characteristics of the patient; and demographic characteristics of the patient.

30 The method may further comprise representing the data as a series of fixed size representations; providing an attention mechanism arranged to identify one or more points of interest within the data based on the fixed sized representations; and providing an output based on the identified points of interest. If a heightened risk of cardiac events is signalled, this output enables a person reviewing the data to identify the regions that led to this signal.

35 Preferably, representing the data as a series of fixed size representations comprises using a network operating over fixed-sized windows of data, for example a neural network and/or a long short-term memory network.

The threshold value is preferably determined based on a plurality of data obtained from multiple sources, where the plurality of data may come from patient data from a plurality of previous patients.

5

The method of analysing cardiac data may be used for, and/or described as, a method of monitoring patients, possibly wherein the method of monitoring patients also comprises a method of indicating a period of increased risk. As aforementioned, patient is intended here to cover a broad range of possible users, so that the method may be used for monitoring a user with a wearable heartbeat monitoring device, even if they are not under observation for any medical conditions (and are considered to be healthy). The method could then, in more general terms, be considered a way of monitoring the health of a user.

Each property has a respective context length. The context length may range between 10 and 100,000 heartbeats, approximately. As will be appreciated, 100,000 beats is roughly the average number of heartbeats a human has in a day, though larger context lengths are possible. More preferably, a context length of around 3,600 beats (i.e. roughly an hour), may be considered. More preferably, a context length of around 350 beats (i.e. roughly 5-6 minutes) may be considered. For example, a context length of around 230 beats has been found to yield good results. Preferably, an optimally discriminating context length is determined, where this determination may be performed using a chi-squared ( $\chi^2$ ) test, a Kolmogorov-Smirnov test, and/or an Energy Test. This context length may be the same for the all properties, more preferably each property has a respective context length (determined specifically for that property).

25

Additional data may be used, where this data is treated similarly to the cardiac data. The method then optionally further comprises: providing further data relating to the patient, wherein the further data comprises at least one of: physiological data, demographic data, admission data, past medical history, laboratory data, imaging data; determining one or more properties of the further data, wherein each property is determined over a particular context length, the context length being selected based on the property; comparing the or each property of the further data against a predetermined threshold value for the further data, thereby to indicate a probability of the patient experiencing a cardiac event; and providing an output based on the comparison. Preferably this output based on the comparison of the further data is combined with the comparison of the (cardiac) data to obtain a combined output.

The data preferably comprises data from multiple heartbeats, more preferably RR intervals of multiple heartbeats, where these are, for example, indicated on an electrocardiogram (ECG). In



order to reduce the processing load, the data may be processed in batches, preferably of at least 5 heartbeats, more preferably of between 5 and 15 heartbeats, yet more preferably of 10 heartbeats. This enables more accurate algorithms to be used with the (batched) data.

5 The properties are preferably properties of multiple heartbeats, such as a mean, a standard deviation, or a standard deviation in successive differences (related to the multiple heartbeats). Optionally, the properties also comprise a measured heart rate variability (HRV), which may be obtained from the RR intervals, and/or a fraction of multiple heartbeats which exceed an abnormality threshold (this may be, for example the fraction of RR intervals related to each  
10 heartbeat which exceed an interval).

Also described herein is a method of analysing cardiac data relating to a patient, comprising: providing cardiac data relating to the patient; determining one or more properties of the data; comparing the or each property against a respective predetermined threshold value, thereby to  
15 indicate a probability of the patient experiencing a cardiac event; and wherein the property comprises a fraction of the multiple heartbeats that exceed an abnormality threshold; and providing an output based on the comparison.

A rate of change of the or each property is optionally determined, where this may be used to  
20 determine an urgency, or may be used to give a further output. This may be separate to any other outputs, or may be combined with other outputs. The method preferably uses a plurality of properties within the data to provide a more accurate output.

If the probability of a cardiac event occurring in a subsequent time period exceeds a certain  
25 threshold, a warning may be issued and the healthcare provider notified. This should enable appropriate action to be taken to prepare for an appropriate response to the cardiac event and/or prevent the arrhythmia from occurring, such as by administering medications or running diagnostic tests. Thus, the predetermined value is determined optimally to separate normal and arrhythmic patients.

30 The probability of a cardiac event is preferably determined using Bayesian inference, where this is used to reduce the number of false positives, or false negatives by creating a link between the predicted (indicated) probability and a prior distribution.

35 Also described herein is a method of analysing data relating to a patient, comprising: providing cardiac data relating to the patient; determining one or more properties of the data; comparing the or each property against a respective predetermined threshold value, thereby to indicate a



probability of the patient experiencing a cardiac event; wherein the indicated probability is calculated using Bayesian inference; and providing an output based on the comparison.

5 There is preferably also presented a measure of the uncertainty related to the indicated probability, where this preferably includes displaying at least one of: a standard deviation, and error bounds. This allows a user to better assess the data as compared to a probability in isolation.

10 The probability is optionally characterised with a measure of the skew of the probability distribution, such as a kurtosis. The probability may be presented as a probability density function (or a cumulative probability function), where this may be displayed graphically, or numerically.

15 This probability is preferably updated periodically, where the time intervals for the updates may depend on the situation. Where the processing volume is not a major consideration, the time intervals may be small enough to be effectively continuous (e.g. they may be less than a second). Where computing power is more of a concern, the time intervals may be at least 5 seconds, or between 10 seconds and 30 seconds, where these values may reduce the computational burden while keeping the risk of missing a cardiac event very low. Where  
20 computing power is a major concern, such as in implanted devices, which have a limited battery, the time intervals may be at least 5 minutes, or at least an hour.

25 Preferably, the time intervals are dependent upon the currently indicated probability, where a very low probability may enable longer time intervals to be used while maintaining a low risk of missing a cardiac event. If this probability begins to increase, the time intervals may be shortened. To further reduce the risk of missing an event, there is preferably a maximum time interval.

30 The probability preferably comprises an indication of a corresponding time, where this may be an amount of time (i.e. a probability is of an event within  $x$  minutes) or one or more time windows (e.g. a probability  $P_1$  of an event between  $x$  and  $y$  minutes, and a probability  $P_2$  of an event between  $y$  and  $z$  minutes). A probability may also display one or more period(s), or time(s), of highest risk, or heightened risk, where this may be related to a probability exceeding a threshold probability of a cardiac event. This may be used to indicate a period over which a  
35 user should be monitored more carefully.

The threshold value against which the property is compared is preferably determined using at least one of: a long short-term memory unit, adversarial training, multi-task training, an

attention mechanism, and a computationally minimalistic algorithm. These techniques may, for example, increase accuracy and/or reduce computational burden.

5 The method optionally also uses an Energy Test to analyse the cardiac data, where the method further comprises: determining an Energy Test metric by performing an Energy test on at least one of the one or more properties of the data; comparing the Energy Test metric to a predetermined threshold value; and presenting an output when the Energy test metric exceeds a predetermined threshold.

10 Also described herein is disclosed a method of analysing cardiac data relating to a patient, comprising: providing cardiac data relating to the patient; determining one or more properties of the data; determining an Energy Test metric by performing an Energy test on at least one of the one or more properties of the data; comparing the Energy Test metric to a predetermined threshold value; and presenting an output when the Energy test metric exceeds a  
15 predetermined threshold.

As has been previously demonstrated in published research, computing features based on RR interval sequences and training classifiers based on these features enables a probabilistic assessment to be made. However, as measurements of individual heartbeats are susceptible  
20 to noise and instrumental fault, a more robust decision-making mechanism is herein provided for deciding whether to issue alerts of a possible oncoming cardiac event.

Optionally, the evaluated property is a property of cardiac data. Preferably, the evaluated property is the RR intervals of multiple heartbeats, for example as indicated on an  
25 electrocardiogram (ECG). Optionally, said value is compared against a predetermined value for a given time window. Optionally, said value is the fraction of said multiple heartbeats that exceed the abnormal threshold (e.g. the number of heartbeats exceeding the abnormal threshold as a fraction of the total number of heartbeats being evaluated).

30 The method may determine the abnormality threshold by training at least two classifiers (which may optionally be cardiac classifiers) to classify a property of multiple heartbeats within the physiological data using at least one machine learning algorithm; and combining the at least two classifiers to produce a hybrid classifier; wherein the combination is based on a performance metric.

35 According to another aspect of the present disclosure, there is disclosed a method of training a hybrid classifier for prediction of cardiac events based on physiological (preferably cardiac) data, the method comprising the steps of: training at least two classifiers to classify a property



of multiple heartbeats within the physiological data using two or more different machine learning algorithms; and combining the at least two classifiers to produce a hybrid classifier; wherein the combination is based on a performance metric.

5 Preferably, training a classifier comprises providing annotated cardiac data, wherein the annotation indicates the occurrence of one or more cardiac events; training a detection classifier to detect cardiac events using the annotated cardiac data; labelling unannotated cardiac data using the trained detection classifier; training a classifier to classify a property of multiple heartbeats using the labelled cardiac data. This method enables a classifier to  
10 be trained using data for which it is not known whether the data relates to an arrhythmia.

Labelling unannotated cardiac data using the trained detection classifier may comprise labelling a subset of unannotated cardiac data dependent upon a threshold probability of correctness. A set of data is optionally split into subsets by the trained detection classifier  
15 based on the probability of the trained detection classifier correctly classifying the data. Those sets of data for which the probability exceeds a threshold are labelled, the rest are not. The trained detection classifier may then be retrained using a training set including the labelled data. The trained dataset may then consider the still unlabelled data, for which the probability of correctly labelling each data set may have increased. This process may  
20 continue, so that a detection classifier is repeatedly retrained and gradually labels an unannotated dataset.

The method may further comprise providing a reference dataset of annotated cardiac data; providing an input dataset of unannotated cardiac data; normalising each member of the  
25 reference dataset and each member of the input dataset to have the same dimensions; comparing each normalised member of the input dataset with one or more normalised members of the reference dataset to identify a measure of similarity; determining labels for the input dataset dependent upon the respective measures of similarity; training a classifier to classify a property of multiple heartbeats using the labelled cardiac data.

30 Preferably, comparing each normalised member of the input dataset with one or more normalised members of the reference dataset comprises determining a root mean square error (RMSE). Preferably, the cardiac data comprises ECG signals.

35 The method may further comprise determining a best performing classifier and a second best performing classifier based upon a performance metric; outputting the classification of



the best performing classifier when the output of the best performing classifier is not close to a decision boundary; and outputting the classification of the second best performing classifier when the output of the best performing classifier is close to the decision boundary.

5

Optionally, the output of the best performing classifier is considered to be not close to the decision boundary when a threshold probability of a correct classification is exceeded.

Each classifier may have a related probability of correct classification for any considered dataset, the output of the best performing classifier is preferably used when the probability of this output being correct exceeds a certain value. The output of the best performing classifier may similarly be considered to be close to the decision boundary when the threshold probability is not exceeded.

The method may comprise training at least two classifiers comprises combining at least two trained classifiers to produce a hybrid classifier; wherein combining the at least two trained classifiers comprises applying weightings to each classifier based on a performance metric associated with each respective classifier.

Preferably, the performance metric comprises at least one of: an accuracy; a sensitivity; a specificity; a precision; and an area under a receiver operating characteristic (ROC) curve.

Optionally, training at least two classifiers comprises using a genetic algorithm and/or simulated annealing. This may improve the performance of the classifier.

As described herein: the abnormality threshold may be determined by: training at least two (preferably cardiac) classifiers to classify a property of multiple heartbeats within said physiological data using at least one machine learning algorithm; and combining said at least two classifiers to produce a hybrid classifier; wherein said combination is based on a metric.

The metric is preferably a performance metric.

Preferably, at least two different machine learning algorithms are preferably used to obtain a result which is less susceptible to flaws within a machine learning algorithm (as different algorithms may be give erroneous results in different situations).

35

Also described herein is a method of training a hybrid classifier for prediction of cardiac events based on physiological data, the method comprising the steps of: training at least two

(preferably cardiac) classifiers to classify a property of multiple heartbeats within said physiological data using one or more machine learning algorithms; and combining at least two classifiers to produce a hybrid classifier; wherein said combination is based on a metric.

5 As mentioned above, the evaluated property may be a property of cardiac data, and preferably the RR intervals of multiple heartbeats. The at least two (i.e. two or more) classifiers may be trained simultaneously. The metric may comprise at least one of: an accuracy; a sensitivity; and a specificity. The one or more machine learning algorithms may comprise at least one of: an Artificial Neural Network; a Support Vector Machines; a k-Nearest Neighbours algorithm; a  
10 Gaussian process; and a Random Forest. All of said classifiers may be trained and combined (and/or added together).

One of the algorithms used optionally comprise a neural network, preferably a convolutional neural network.

15 A distilling method is preferably used within the training of the classifiers, where this comprises: training a first neural network; and training a second neural network dependent upon the output of the first neural network. Distilling, as described, may be used to train a second, simpler and faster, model from a first model.

20 When using a Gaussian process, preferably a limited number of datapoints are used, preferably where this limited number is no more than 3,000. This has been found to result in an improved classifier.

25 The hybrid classifier may be further configured to output a weighted sum of the outputs of the at least two classifiers. The weights in the weighted sum may be determined from a measure of performance of each of the classifiers in the hybrid classifier.

30 The root mean square error (RMSE) may be used to determine an optimal combination of classifiers. This RMSE is preferably the RMSE over misclassifications.

As aforementioned, testing time domain measures for optimal context may be determined by performing a chi-squared ( $\chi^2$ ) test of compatibility on different context lengths. Context lengths from 10 to 100,000 beats may be considered during the chi-square ( $\chi^2$ ) test. As mentioned  
35 above, preferably a context length of under an hour, or around 3,600 beats, may be considered. More preferably, a context length of around 350 beats may be considered. For example, a context length of around 230 beats has been found to yield good results.



Maximally discriminating lengths may be determined for each feature individually prior to training the classifiers, thereby to achieve enhanced separation power.

5 There are multiple ways to measure the performance of a classification task. Commonly used metrics include accuracy, sensitivity, specificity, F-score, and precision. However, optimising the hyper-parameters of an algorithm based on any one of these metrics leads to suboptimal performance overall. In the method described herein, a custom proper score is employed to achieve maximally discriminating results. Specifically, the RMSE evaluated for misclassifications is minimised at multiple stages of the method including at the neural network training, heartbeat-level and patient-level separation steps.

15 In the method described herein, multiple heartbeats, each with corresponding contexts, may be evaluated before a decision to issue an alert is made. The fraction of heartbeats that exceed an abnormality threshold is computed. Alerts are only issued after the fraction has been evaluated in an appropriate time window and found to exceed a value that optimally separates normal and arrhythmic patient groups.

20 A hybrid classifier leverages the strength of each method and results in more robust performance. In the method described herein the root mean square error (RMSE) is employed to arrive at an optimal combination of classifiers.

25 The method described herein tests time domain measures for optimal context. Context lengths from 10 to 100,000 beats (more preferably context lengths from 10 to 3,600), for example, context lengths from 10 to 350 beats, are considered and a chi-square test of compatibility is performed. Maximally discriminating lengths are determined for each feature individually prior to training the classifiers thus achieving enhanced separation power. Previously, a default five-minute window was used in computation of features, based on a qualitative understanding of the heart and commonly used heuristics in the field. Until now, however, no attempt has been made to determine if a five minute window is actually appropriate for all features and/or if optimal discriminating power is achieved for all variables with such a predetermined time window.

Also disclosed herein is a method of monitoring patients using any of the above methods.

35 Also described herein is a system for predicting cardiac events using physiological data relating to a patient, comprising: means for providing physiological data relating to the patient; and an analysis module configured to evaluate a property of multiple heartbeats in the physiological data, determine whether said property exceeds an abnormality threshold and derive a



probability of the patient experiencing a cardiac event; and means for providing an output; wherein the analysis module is configured to trigger an output if the probability of the patient experiencing a cardiac event in the subsequent time period is greater than a predefined threshold.

5

Also described herein is a system for analysing cardiac data relating to a patient, comprising: means for providing cardiac data relating to the patient; an analysis module for determining one or more properties of the data, wherein each property is determined over a particular context length, the context length being selected based on the property; a comparison module for  
10 comparing the property against a predetermined threshold value, thereby to indicate a probability of the patient experiencing a cardiac event; and a presentation module for providing an output based on the comparison.

15

According to yet another aspect of the disclosure herein, there is described a system for analysing cardiac data relating to a patient, comprising: means for providing cardiac data relating to the patient; an analysis module for determining a property of the data, wherein the property is determined over a particular context length, the context length being selected based on the property; a comparison module for comparing the property against a predetermined threshold value, thereby to indicate a probability of the patient experiencing a cardiac event;  
20 and a presentation module for providing an output based on the comparison.

20

The analysis module may comprise a hybrid classifier trained according to the methods described above and herein.

25

Optionally, the means for providing cardiac data relating to the patient comprises a spatially separated measurement module. This may be used, for example, where a patient is using an implantable device. Data recorded by this device may be provided to a (spatially separate) server, where it is analysed. An output may then be displayed to the user and/or another person (such as a doctor).

30

Also described herein is a client terminal connectable to the system disclosed, where this may be used to access the output in a format desirable to the user. This may, for example, be a handheld portable device which a user could use to connect to a server which performs the methods disclosed herein. Such a server is also disclosed herein.

35

The physiological (preferably cardiac) data may be provided by / obtained (e.g. sourced) from at least one of: an electrocardiogram (ECG) machine; a pulsometer; a wearable cardioverter defibrillator; an implantable cardioverter defibrillator; a respiratory monitor; and a capnography

monitor, or other such source extracting data from the cardiorespiratory system of a patient. The analysis module may comprise a hybrid classifier trained as described above and herein.

5 Also described herein a portable and/or wearable device, which is configured to carry out the disclosed methods.

Also described herein is a machine learning algorithm for predicting cardiac events. The invention extends to a method and/or a system for predicting a cardiac event substantially as described herein and/or as illustrated in the accompanying figures.

10 From a physiological standpoint, the method described herein is configured to probe the behaviour of the autonomous nervous system by measuring heart rate variability (HRV) using physiological data. Optionally, respiratory rate variability may be used in addition to cardiac data. Further features may be developed based on respiratory rate variability and included as  
15 additional input to the classifiers during training.

The method described herein was developed using a number of datasets containing cardiac data, e.g. the “Spontaneous Ventricular Tachyarrhythmia (VTA) Database”, obtained from Medtronic, Inc. (<http://www.physionet.org/physiobank/database/mytdb/>). In particular, the  
20 method described herein was developed using RR intervals (based on the above-mentioned datasets) that enable heart rate variability (HRV) analysis to be performed. Other information present on an ECG or extracted from other monitoring devices may also be incorporated, and, for example, respiratory data may further be added as an input.

25 As used herein, the term “RR interval” preferably refers to an interval from the peak of one “QRS complex” to the peak of the next “QRS complex” (i.e. the time interval between two consecutive “R waves”) seen on an electrocardiogram (ECG). The RR interval may be used to assess the ventricular rate. Such data can also be extracted from pulsometers. As used herein, the term “QRS complex” preferably refers to a combination of three graphical deflections seen  
30 on an ECG, which is usually the central and most visually obvious part of the tracing on the ECG. It corresponds to the depolarization of the right and left ventricles of the human heart. Both features can be seen on the exemplary ECG illustrated in Figure 1.

As used herein, the term “cardiac event” preferably connotes a change in the cardiac rhythm of  
35 a patient, for example from normal sinus rhythm to an arrhythmia; from one type of arrhythmia to another; or a change in the severity or dangerousness of a cardiac rhythm. In particular, cardiac events may refer to changes which may cause sudden cardiac death (SCD), such as the occurrence of ventricular tachyarrhythmias (VTA).



As used herein, except where the context requires otherwise, the term "comprise" and variations of the term, such as "comprising", "comprises" and "comprised", are not intended to exclude further additives, components, integers or steps.

5

The invention extends to methods, system and apparatus substantially as herein described and/or as illustrated with reference to the accompanying figures.

10

The invention also provides a computer program or a computer program product for carrying out any of the methods described herein, and/or for embodying any of the apparatus features described herein, and a computer readable medium having stored thereon a program for carrying out any of the methods described herein and/or for embodying any of the apparatus features described herein.

15

The invention also provides a signal embodying a computer program or a computer program product for carrying out any of the methods described herein, and/or for embodying any of the apparatus features described herein, a method of transmitting such a signal, and a computer product having an operating system which supports a computer program for carrying out the methods described herein and/or for embodying any of the apparatus features described

20

herein. The invention also provides a computer program and a computer program product comprising software code adapted, when executed on a data processing apparatus, to perform any of the methods described herein, including any or all of their component steps.

25

The invention also provides a computer readable medium having stored thereon the computer program as aforesaid.

30

Other aspects of this system, client device and/or method may be implemented in software running on various interconnected servers, and it is to be appreciated that inventive aspects may therefore reside in the software running on such servers.

35

The invention also extends to a server or a plurality of interconnected servers running software adapted to implement the system or method as herein described.

Any feature in one aspect of the invention may be applied to other aspects of the invention, in any appropriate combination. In particular, method aspects may be applied to apparatus aspects, and vice versa.



Furthermore, features implanted in hardware may generally be implemented in software, and vice versa. Any reference to software and hardware features herein should be construed accordingly.

5

Any apparatus feature as described herein may also be provided as a method feature, and vice versa. Furthermore, as used herein, means plus function features may be expressed alternatively in terms of their corresponding structure.

10 It will be appreciated that particular combinations of the various features described and defined in any aspects can be implemented and/or supplied and/or used independently. In other words, any feature in a particular aspect may be provided independently and/or applied to other aspects, in any appropriate combination.

15 At least one exemplary embodiment of the present invention will now be described with reference to the accompanying figures, wherein similar reference numerals may be used to refer to similar features, and in which:

Figure 1 shows an example of a typical electrocardiographic tracing;

Figure 2a is a general process flowchart for the method described herein;

20 Figures 2b and 2c are a specific process flowchart for the method described herein as pertaining to one of the possible inputs to the algorithm;

Figures 3a and 3b show all of the heartbeats in a "Spontaneous Ventricular Tachyarrhythmia (VTA) Database", before and after outliers have been removed, respectively;

25 Figures 4a and 4b show distributions of RR intervals for a heart, specifically the distribution preceding an arrhythmia (circles), the distribution 5 minutes prior to the arrhythmia (triangles) and the distribution for a normally functioning heart (squares)];

Figure 5 shows the time evolution of the mean RR interval leading up to an arrhythmia and the time evolution for a normally functioning heart;

30 Figure 6 shows the time evolution of one of the time domain inputs to the algorithm, the standard deviation in RR intervals;

Figures 7a and 7b show the statistical compatibility of the SD1 and SD2 variables between the 'arrhythmic' and 'normal' distributions as a function of a context length;

Figures 8a and 8b show probability density functions for the standard deviation of functions modelling a distribution of heartbeats;

35 Figures 9a and 9b show probability density functions for the fraction of ectopic beats of a distribution of heartbeats;

Figure 10 shows a flowchart for a method of training a classifier to use unannotated data;

Figure 11 shows a flowchart for a method of identifying cases of ventricular tachycardia;

Figure 12 shows the time evolution of the probability for an arrhythmic episode for the Random Forest classifier;

5 Figure 13 shows the separation between the 'arrhythmic' and 'normal' probability distributions as a function of probability in units of standard deviations for the Random Forest classifier;

Figure 14 shows the distribution of fraction of abnormal beats for 'arrhythmic' and 'normal' patients scaled to unit area such that the y-axis scale is in arbitrary units (A.U.);

10 Figures 15a – f show performance metrics for a boosted decision tree classifier and a committee voter classifier;

Figure 16 shows an exemplary system for predicting cardiac events; and

Figure 17 shows a component diagram for analysing patient data and displaying an output; and

Figures 18a and 18b show an output that indicates an increased risk of a cardiac event.

15 In what follows, cardiac data that terminate with a Ventricular Tachyarrhythmia (VTA) is referred to as 'arrhythmic', and cardiac data from control samples is labelled as 'normal'.

#### Prediction method for use on patients

20 Figure 2a illustrates a method of predicting cardiac events. Physiological data, obtained from a monitoring device are input into an analysis module comprising a pre-trained classifier. The physiological data can comprise data relating to the patient that is collected in real time, for example cardiac data. In some embodiments, the physiological data alternatively or additionally  
25 comprises respiratory data relating to the patient collected from a respiratory monitoring device, which are also input into the pre-trained classifier.

The analysis module uses the input physiological data to analyse the heartbeat of the patient, and determine one or more probabilities of the patient experiencing a cardiac event within a  
30 period of time in the future.

RR interval sequences, as illustrated in Figure 1, are taken as input, data analysis is performed, and a classifier separates 'arrhythmic' and 'normal' beat sequences. The classifier attributes a probability to each heartbeat and then aggregates the output into an abnormality  
35 fraction in [0,1] that forms the basis for a decision to alert the healthcare provider. The abnormality fraction may thereby serve as a useful, actionable, easy-to-interpret number that



may guide healthcare providers, patients, or other people who may be in a position to assist the patient. The warning may enable preparation for an appropriate response to the cardiac event and/or prevention of the cardiac event, such as by administering medications and running diagnostic tests.

5

The method will now be described in more detail, with reference to the process flow illustrated in Figures 2b and 2c. Physiological data, in this example in the form of RR intervals, are input into the analysis module from a monitoring device. Physiological data may contain false measurements (e.g. “outliers”) owing, for example, to movement of the patient and/or poor connections in the monitoring device, which lead to artefacts in the datasets.

10

Patients that suffer from VTAs are also likely to suffer from ectopic beats such as premature ventricular complexes. Therefore, as indicated in the process flow in Figure 2b, it is necessary to identify and remove any outliers from the physiological data. This may be achieved using, for example, criteria described in G. D. Clifford, F. Azuaje, and P. E. McSharry, *Advanced Methods and Tools for ECG 223 Data Analysis*, Artech House Publishers, 2006. The presence of ectopic beats is, however, recorded and used in the subsequent analysis (see below).

15

20

The effect of outlier removal on the data is illustrated in Figures 3a and 3b. Figure 3a illustrates the raw RR interval data taken from the "Spontaneous Ventricular Tachyarrhythmia (VTA) Database", which comprises multiple outlying datapoints 10. The cleaned version of the same data is shown in Figure 3b.

25

The cleaned physiological data are then pre-processed as indicated in the process flow of Figures 2b and 2c to obtain unbiased measurements of frequency domain parameters (as discussed below). More specifically, the data first undergoes cubic spline interpolation, and is then resampled at, for example, 7Hz. Subsequently, the spectral power is computed using a Welch periodogram, for example with a 256-point window overlapped at 50%.

30

A series of derived quantities are computed based on RR interval data. The derived quantities (listed below) are referred to (interchangeably) as ‘features’ or ‘properties’:

*i) Time Domain*

35

- The arithmetic mean,  $\mu$  of the RR intervals;
- The standard deviation,  $\sigma$  of the RR intervals;
- The standard deviation in successive differences,  $\sigma_{\text{Diff}}$ , of the RR intervals.



The distribution of RR intervals in the time domain can provide valuable data relating to the probability of a patient undergoing a cardiac event.

5 For example, Figures 4a and 4b illustrate distributions of RR intervals for Arrhythmic and Normal sets of heartbeats, measured in arbitrary units (A.U.). The distribution of normal heartbeats is shown using squares, while that of the arrhythmic heartbeats is shown using circles. Furthermore, the distribution of arrhythmic heartbeats with five minutes of data prior to the arrhythmia is shown using triangles.

10 Figure 5 shows the time evolution of the mean RR interval leading up to an arrhythmia at  $t=0$  s for the 'Arrhythmic' distribution. In particular, Figure 5 illustrates the dramatic drop in mean RR intervals at the onset of the arrhythmia near  $t=0$  s.

15 Figure 6 shows the time evolution of one of the time domain inputs to the algorithm, the standard deviation in RR intervals. The cardiac event occurs at  $t=0$  s, which appears at the leftmost point of the x-axis. Time flows from right (the past) to left (terminating at the event).

#### *ii) Nonlinear Poincaré*

- Poincaré nonlinear analysis variables,  $SD1$ ,  $SD2$ , and  $SD1 / SD2$ .

20 A Poincaré HRV plot is a graph in which successive RR intervals are plotted against one another. From this plot values for  $SD1$  (the dispersion of points perpendicular to the line of identity) and  $SD2$  (the dispersion of points parallel to the line of identity) are determinable. These plots, and the determination of the  $SD1$  and  $SD2$  values, are well known.  $SD1$ ,  $SD2$ , or  
25 a combination of  $SD1$  and  $SD2$  are used as inputs to the AI classifier.

#### *iii) Sample Entropy*

- Sample entropy over four epochs,  $S1$ ,  $S2$ ,  $S3$  and  $S4$ .

#### *iv) Frequency Domain*

- Frequency domain parameters,  $VLF$ ,  $LF$ ,  $HF$  and  $LF/HF$ , derived from the spectral power calculated from the Welch periodogram.

#### *v) Ectopic Beat Frequency*

- 35
- The relative frequency of ectopic beats,  $f_e$ .

The optimal context for each feature, *i.e.* the optimal – or maximally discriminating – ‘context length’ (as discussed below) for determining whether a feature is indicative of a cardiac event, is determined before each feature is input into an Artificial Intelligence based classifier.

5 The features derived from the RR interval data are input into an Artificial Intelligence Based Classifier (the AI classifier). The AI classifier can comprise a pre-trained classifier, or preferably multiple pre-trained classifiers combined into a hybrid classifier, that has been trained (as described below) to identify abnormal beats in the physiological data by assigning a probability (i.e. a number in [0,1]) to each heartbeat that reflects the likelihood for the given heartbeat to  
10 lead to an arrhythmic episode.

In some embodiments, in order to arrive at a robust decision, the number of ‘abnormal’ heartbeats (e.g. which cross a threshold probability) are counted, and the fraction of said ‘abnormal’ heartbeats occurring in a given time window (for example, five minutes) is  
15 computed. This leads to an abnormality fraction,  $F$ , which is attributed to each patient. A ‘yes/no’ decision is then made based on this fraction, and an alert may be issued (or another action taken) for positive decisions. The alert may, for example, indicate that a cardiac event is predicted; in some embodiments, it also provides additional data related to the probability of the event occurring.

20 The counting of ‘abnormal’ heartbeats may also be used to obtain a rate of change of the occurrence of ‘abnormal’ heartbeats, where this rate of change may be used to identify both that a cardiac event is likely, and also to predict an urgency – where a high rate of change may indicate that a cardiac event is likely to occur soon.

25 More specifically, as is shown in Figure 2c, the AI classifier comprises a beat-level classifier, a patient-level classifier, and a decision-level classifier.

The beat-level classifier is trained to identify abnormal heartbeats within a dataset of  
30 heartbeats. These data, and the output of the beat-level classifier, is combined with patient-level data within the patient-level classifier, within which the beat data is assessed along with contextual patient data, such as an Electronic Health Record, or a record of existing health conditions, to determine whether the beat-level data is indicative of an arrhythmic episode in a particular patient, or set of patients. The output of the patient-level classifier is fed into the  
35 decision-level classifier, which is trained to combine data from the preceding classifiers to output the probability of the given heartbeat indicating an upcoming arrhythmic episode. The decision-level classifier may be optimised for a certain metric, e.g. accuracy or specificity, as is described below.



The training of these classifiers occurs on three levels corresponding to the classifiers themselves: the beat-level, the patient-level, and finally the decision-level. At the beat-level, classifiers are trained to separate arrhythmic and normal heartbeats. At the patient-level, classifiers are trained to separate arrhythmic and normal patients with a combination of beat-level and patient-level inputs.

In practice, this typically involves data first being examined on the heartbeat-level and then, if the beat-level classifier indicates the data as high risk, data is further examined on a patient level. By using a multi-level system, the classifier is capable of accounting for contextual data, such as a patient having an abnormally high resting heart rate.

Examples of patient-level inputs include the arithmetic mean, the standard deviation, and an abnormality fraction computed from beat-level classifier outputs. Other aspects of the patient, such as those conventionally found in an Electronic Health Record, can also be incorporated at this stage. At the decision-level, the entire process is optimised for a metric such as accuracy or specificity by scanning over the space of all classifier hyperparameters. In some embodiments, a short long-term memory model is used and the hyperparameters comprise the learning rate and/or the network size).

In some embodiments, the hyperparameters used within one or more of the classifiers are optimised using evolutionary algorithms, preferably genetic algorithms. Characteristics of the classifiers are modified in a random or semi-random manner and the resulting performance of the classifiers is compared to the non-modified classifier. This process is repeatedly performed for a number of modified classifier architectures (those showing potential improvement over the non-modified classifier); this may lead to the discovery of well-performing hyperparameters that would not otherwise be considered.

In some embodiments, simulated annealing is used in order to optimise at least one of the beat-level classifier, patient-level classifier, and decision-level classifier.

### Classifier training/architecture

The AI classifier, and more specifically each of the beat-level classifier, patient-level classifier, and decision-level classifier, can be trained by a machine learning system receiving as input examples of heartbeats from a training dataset comprising known normal and abnormal heartbeats from which the system can learn to identify whether an arrhythmia is likely to occur. Each heartbeat in the training data set is represented as a real-valued vector containing values



for features that describe the specific heartbeat, and enable a classification to be made. The training data is pre-processed in the same way as described above in relation to Figure 2b, providing the same features that will be used in the prediction method for use in the training process.

5

Each of the beat-level classifier, the patient-level classifier, and the decision-level classifier may be trained using any combination of the methods described below.

10

There is freedom in the number of preceding heartbeats that should be included in the computation of a feature. This is referred to herein as 'context length'. Multiple context lengths from 10 beats to 100,000 beats (though preferably context lengths of less than around 3,600) are considered as variables for time domain measures ( $\mu$ ,  $\sigma$ , and  $\sigma_{\text{Diff}}$ ) and Poincaré nonlinear analysis.

15

A  $\chi^2$ -test ('chi-squared' test) for statistical compatibility is performed for each 'feature' (i.e. derived quantity) and each context length between the 'arrhythmic' and 'normal' data sample distributions. Context lengths that are optimally discriminating, *i.e.* where the data range is the most significant for detecting a cardiac event, can then be selected as evidenced by a large  $\chi^2 / ndf$  between the respective distributions, where "*ndf*" is the number of degrees of freedom.

20

Referring to Figure 7a, there is shown for the SD1 variable the variation of the  $\chi^2$  variable against the number of beats (context length) for a distribution of heartbeats. The data has a consistent number of degrees of freedom (so that for the whole of this dataset  $\chi^2$  is proportional to  $\chi^2 / ndf$ ). From Figure 7a it can be seen that a context length of 190 beats is optimally discriminating for the SD1 variable for this distribution.

25

Referring to Figure 7b, there is shown the variation of the  $\chi^2 / ndf$  variable against context length for the SD2 variable. It can be seen that for the same distribution of heartbeats the optimally discriminating context length for the SD2 variable is 230 heartbeats,

30

In some embodiments, the input datasets are processed to reduce the effect of statistical fluctuations present in the histograms of the properties (e.g. SD1, SD2). This reduces the effects of the binning density chosen to analyse those histograms.

35

In some embodiments, an adaptive kernel density estimation technique is used in which the histograms are smoothed with a function, preferably a continuous function, that represents the distribution of the property. Typically, a probability density function (PDF) is used for this smoothing, where an appropriate PDF may be determined by superposing Gaussian

distributions with equal surface, but varying width. The width of each Gaussian is dependent upon the local event density of the measured histogram; generally, a wide Gaussian distribution is used if the local event density is low and a narrow Gaussian distribution is used if the local event density is high.

5

In typical embodiments, the primary feature used to discriminate between normal and arrhythmic distributions is the mean of the Gaussian distributions determined as suitable for smoothing the histogram of a considered property. In some embodiments, the standard deviation of the determined Gaussian distributions is also considered. The Gaussian distributions used for smoothing a measured dataset are compared to predetermined distributions for measuring known normal and arrhythmic cardiac datasets, this is useable to evaluate an input dataset.

10

In practice, determining a PDF typically involves monitoring cardiac data over a period of time. Using the collected data, a distribution can be determined that is suitable for representation as a histogram. Using the measured distribution, a PDF that approximates the distribution of the property is determined; this PDF approximates the shape of an unbinned feature distribution and so reduces the adverse effects from sub-optimal binning.

15

As an example, and for illustrative purposes only, consider a dataset containing one point for an SD1 of 90 and one point for an SD1 of 92. Since the variance within this range is more likely due to a lack of data than a large variance within the probabilities of the considered SD1s, it might be more appropriate to use a bin of size five from 90 – 95 than five bins of size one. Were a bin size of one to be used, smoothing the resultant histogram using a PDF would reduce the effect of the binning size, thereby avoiding potentially confusing peaks.

20

25

In order to obtain PDFs to compare the measured data to, PDFs are predetermined for normal and arrhythmic datasets. This enables (the features of) a PDF determined using measured data to be evaluated (e.g. compared to threshold values that are indicative of arrhythmia). More specifically, descriptors of the PDFs used, such as the mean, variance, and/or kurtosis, are used within a comparison.

30

PDFs may be determined for specific situations, for example there may be determined a PDF for arrhythmic heartbeat data in patients over 60 with pre-existing heart conditions. The threshold values used for comparing descriptors of a determined PDF are then selected from an appropriate predetermined PDF.

35



Each PDF is scaled to enable comparison between datasets. This comprises scaling the PDFs to the number of events and/or scaling the histogram of event data so that it is suitably represented by a PDF. As an example, the histogram of each dataset may be modified where each bin is divided by the number of total events, to obtain a binned probability measurement.

5 This may then be modelled by a PDF to reduce the effect of the chosen binning size.

Figures 8a and 8b illustrate PDFs for the standard deviations of Gaussian distributions that represent normal and arrhythmic heartbeats for context lengths of 20 and 290. From these figures, it can be seen that using either context length, arrhythmic distributions can be distinguished from normal distributions. With large context lengths, the probability density functions of the standard deviation are “stretched” making this assessment simpler.

Figures 9a and 9b show PDFs for the ectopic beat fractions for context lengths of 10 and 310. Again, it can be seen here that the longer context length achieves better discriminating power; this is shown by the ‘shoulder’ present for fractions between 0.1 and 0.4 when a context length of 310 heartbeats is used.

More generally (for PDFs and for other analysis methods), there is an optimal context length determined above which the discriminating power decreases or does not increase substantially. This optimal context length is that which is typically used for analysis. Determining the optimal context length for each feature preferably occurs prior to training. The context length is then held constant during the classifier training phase.

A maximum context length may also be enforced in order to limit the data storage needed, the recording time needed, and to ensure that a rapid decision is possible. The 3,600 beats mentioned previously may be used to limit the amount of data which must be considered.

In order to use the available dataset maximally, a 10-fold cross-validation is performed, whereby the dataset is divided into ten parts and the model is trained ten times. Each time, eight parts are used for training, one part for hyper-parameter tuning and one part for testing. The assignment of different folds is rotated during the ten times.

Five separate machine learning algorithms, in particular, can be used in order to train classifiers (although this method is, of course, extendable to other algorithms). The algorithms are then, preferably, later combined to form a hybrid algorithm, in order to take advantage of each of their strengths.



In some embodiments, the combining of algorithms comprises “committee voting”, where the outputs from each classifier are combined, with these outputs weighted dependent upon the performance of the corresponding classifier. Better performing classifiers, as determined using the metrics described below, are given higher weightings and therefore have a larger effect in the determination of the committee classifier output.

In some embodiments, “ask a friend” voting is used, where the best performing classifier is used, unless the output of the best performing classifier is close to the decision boundary associated with this classifier. The decision boundary is a boundary, as described with reference to equation 1.1, that separates distributions identified as normal from distributions identified as arrhythmic; close to a decision boundary there is an increased possibility of obtaining an incorrect output (e.g. a false positive or a false negative) from the classifier associated with this decision boundary. Therefore, the best performing classifier is used where it achieves an output far from the associated decision boundary, the second best performing classifier is used either instead of or in combination with the best performing classifier when the best performing classifier obtains an output close to this decision boundary.

In some embodiments, being close to the decision boundary relates to being beneath a certain threshold probability of correctness, where the probability of correctness for a given output is determined by considering the characteristics of the classifier, e.g. the classifier may output a probability of correctness.

In some embodiments, both “committee voting” and “ask a friend” voting are used, where a “committee voting” classifier is formed from the weighted classifiers and this classifier is considered within the “ask a friend” voting.

In some embodiments, the classifier is a long short-term memory unit which may record values over an arbitrary time interval. This type of classifier is particularly useful for processes which have time lags between events (such as cardiac events).

In some embodiments, a convolutional neural network could be used to detect patterns within the recorded data, where this may be combined with an attention mechanism. An attention mechanism enables the neural network to ‘learn’ where it needs to focus and dynamically assign more importance to those areas. The attention mechanism calculates a weight for each time-window in the input stream and uses it to scale the importance of information coming from that window. This method has been shown to be very successful in other domains such as language processing and also enables visualisation of where the model is focusing, thereby making the actions of the system more human-interpretable.

More specifically, in some embodiments, the neural model is arranged to represent the temporal stream as a series of fixed sized representations. This can be achieved using a long short-term memory (LSTM) architecture or a convolution neural network operating over fixed-sized windows. An attention mechanism is constructed on top of the network to allow the model to dynamically predict how much focus should be assigned to each position in the temporal stream. When analysing the cases where the model predicts positive labels for VTA, the attention weights are usable to visualise which areas in the signal were most important for making the prediction. In addition, the gradient on individual feature vectors is usable to find which specific features were most important for making the prediction at that time. This allows specific features and specific time in the data stream to be flagged as of relevance, which enables a practitioner to rapidly identify relevant parts of a data stream. Therefore an informed decision can be made regarding the health of the patient, furthermore anomalies within the data that are worthy of further inspection are observable.

15

### 1. *Artificial Neural Network*

The feature vectors are given as input to an artificial neural network consisting of three layers. The first layer is an “input layer”, the size of which depends on the number of features in the feature vectors. The second layer is a “hidden layer” with *tanh* activation, with size 10. Finally, the third layer is a single neuron with sigmoid activation. The neurons in the hidden layer will automatically discover useful features from the input data. The model can then make a prediction based on this higher-level representation. The network may be optimised using AdaDelta, for example. Parameters may be updated based on mean squared error as the loss function. The model may be tested on the development set after every full pass through the training data, preferably wherein the best model is used for final evaluation.

25

### 2. *Support Vector Machines (SVM)*

Support Vector Machines (SVM) are a separate class of supervised machine learning algorithms. Instead of focusing on finding useful features, they treat the problem as a task of separation in a high-dimensional space. Given that the feature vectors contain  $n$  features, they aim to find an  $n-1$  dimensional hyperplane that best separates the positive and negative cases. This hyperplane is optimised during training so that the distance to the nearest datapoint in either class is maximal.

30

### 3. *k-Nearest Neighbours*

35

$k$ -Nearest Neighbours ( $k$ -NN) is an algorithm that analyses individual points in the high-dimensional feature space. Given a new feature vector that we wish to classify,  $k$ -NN returns  $k$  most similar points from the training data. Since we know the labels of these points,  $k$ -NN



assigns the most frequent label as the prediction for the new point. This offers an alternative view to the problem - it no longer assumes that heartbeats of a single class are in a similar area in the feature space, but instead allows us to look for individual points that have very similar features.

5

#### 4. *Gaussian Process*

Gaussian Process is a statistical model where each datapoint is associated with a normally distributed random variable. The Gaussian Process itself is a distribution over distributions, which is learned during training. This model associates each prediction also with a measure of uncertainty, allowing us to evaluate how confident the model is in its own classification. As this type of model is difficult to train with more than 3,000 datapoints, it is preferable to ensure that a suitable size is sampled during training.

10

#### 5. *Random Forest*

Random forests are based on constructing multiple decision trees and averaging the results. Each decision tree is a model that attempts to separate two samples based on sequential splittings for each input feature. In this implementation, datapoints that are misclassified are given a weight larger than one (referred to as 'boosting' or as a 'boosted decision tree' method).

15

20

Each classifier assigns a probability (i.e. a number in  $[0,1]$ ) to each heartbeat that reflects the likelihood for the given heartbeat to lead to an arrhythmic episode. Several different thresholds for the probability may be considered and the value that optimally separates the 'arrhythmic' and 'normal' datasets is chosen. This may be referred to as optimal classification separation.

25

30

In some embodiments, the methods of predicting cardiac events are used (and/or embedded) within a portable device, such as a pacemaker, or an implantable cardioverter-defibrillator. Within such a device, it is important that computations are minimised, to maximise the battery life of the device. In order to achieve this algorithms with low computational cost are used (possibly at the expense of some accuracy).

35

An example of using low computational cost algorithms is the use of difference of area (DOA) methods, which have a low complexity, within waveform analysis. Bin area methods (BAM) may also be used as these provide a trade-off between complexity and accuracy. More generally, it is preferable to use algorithms which analyse time domain features as opposed to those which analyse frequency domain features.



In order to speed up the execution of the Random Forest algorithm, in some embodiments each input feature is discretised so that the volume of information fed to the decision trees is reduced. This approach is used to speed up the execution of the classifier and to reduce the effect of noise by choosing step sizes greater than the fluctuations present in the features on account of noise.

In some embodiments, classifiers are formed using 'distilling'. First, a very complex and computation-intensive neural network is trained. Next, a simpler and faster model is constructed, before being trained it on the output of the former model. This approach results in models (and classifiers) that have the benefits of both speed and accuracy.

'Batching' is another method that is used in some embodiments to speed up computation. If a model has limited processing power and cannot process one heartbeat at a time, the incoming data can be combined into batches of ten heartbeats to reduce the computational burden. This results in the model being up to ten beats behind in making predictions, but enables the use of more accurate models.

In some embodiments, an adversarial training model is used, where cases for which the classifier would misclassify data are determined and these cases are used to improve the performance of the classifiers.

As an example: a neural network is provided that is trained to classify RR sequences. Starting with a healthy rhythm, it is determined which (small) changes need to be made to this rhythm in order for the network to misclassify it as a VT example. This method then enables identification of the weak points of the network. These examples (of misclassified datasets) are subsequently introduced into the training data and the classifiers are trained to classify them correctly. This results in a more robust model with a decreased likelihood of misclassifications.

In some embodiments, existing training data, include small random noise in the signal, is added to the training set with the same labels. Given that the noise is small, it is valid to expect that the true label of these examples should not change. Using these data for training introduces the model to a wider variation of datapoints around the known recorded instances, making it more robust during testing.

In some embodiments, noise is generated such that it maximally confuses the model. Using gradient descent, it is calculable how individual feature values should be changed in order to make the model give a wrong prediction. L2 regularization may be used to ensure that the modifications will be minimal, therefore the true label of the example can be assumed to be the

same as the original datapoint and any mistakes the model makes are due to discovered blindspots.

5 By including these adversarial examples in the training data, the model is able to learn to correct for these incorrect predictions. The process can be repeated iteratively to continue improving the model.

10 Annotated data for VT/VF detection and prediction is very limited and therefore it is beneficial to make use of available data in unannotated datasets of ECG signals. In some embodiments, a detection system is trained and using it to return new examples for training the prediction system. Referring to Figure 10, such a method is shown.

15 In a first step 102, the available annotated data is used to train a detection system. This training uses machine learning techniques as are known and/or as are described herein.

In a second step 104, the detection system is applied on unannotated data.

20 In a third step 106, the detection system returns a labelled subset of the data that the model finds most likely to be positive (e.g. indicative of there being a cardiac event) and/or a labelled subset of the data examples that the model finds most likely to be negative (e.g. indicative of there not being a cardiac event).

25 In some embodiments, the detection system labels a subset of data for which the probability of a correct label being applied, as determined by the detection system, exceeds a certain threshold. This enables the provision of a large training set while minimising the risk of incorrect labelling.

30 In a fourth step 108, an updated training set is provided comprising the original annotated data and the labelled data. While the labels are not guaranteed to be correct, they are likely to assist the prediction system – that used to predict cardiac events in patients – and move the performance closer to the detection system.

35 In a fifth step 110, both the detection system and the prediction system are retrained with the updated training set.

In a sixth step 112, it is determined whether the performance of the prediction system is improving on a dedicated test set.



If the performance of the prediction system is improving, the process is repeated from the first step 102. That is, the new data is used as annotated data within the training of the detection system, and this retrained detection system is used to label a subset of the remaining unannotated data. With each iteration, the retrained detection system is better able to label the previously unlabelled data, so that a repeatedly retrained detection system is able to piecemeal label an unannotated dataset with only a small possibility of erroneous labelling.

Once the performance of the prediction system is no longer improving, in a seventh step 114, the detection system and the prediction system are output.

In some embodiments, repeating the process involves reclassifying data within the third step to determine whether there is an improvement in the performance of the prediction system, e.g. a datapoint previously labelled as positive is labelled as negative and it is determined whether this relabelling improves the performance of the prediction system. In some embodiments, the measure of improvement may be improvement within a specified number of iterations, so that a single iteration without improvement does not halt the method.

Similarly, the number of available training examples for Ventricular Tachycardia identification is limited. In some embodiments, there is performed a method for VT detection within unlabelled datasets based only on a subset of known examples and the presence of wide QRS complexes, which is an identifying feature of Ventricular Tachycardia. Such a method is shown in Figure 11.

In a first step 122, individual QRS complexes are extracted from both the reference ECG signal (the known examples) and an input ECG signal (unknown data). Existing algorithms for RR interval conversion can be used for this.

In a second step 124, the complexes are normalised into the same range in each dimension so that they are comparable.

In a third step 126, the similarity between the shapes of the complexes from both the input and reference signals is calculated. Root mean square error (RMSE) is used for comparing the shapes.

In a fourth step 128, it is determined whether the similarity is higher than an assigned threshold.

If the similarity is higher than the threshold, in a fifth step it is determined that, based upon the exceeding of the threshold, the QRS complex is abnormally wide and should therefore be assigned for manual review or directly classified as Ventricular Tachycardia.

5 In various embodiments, the similarity threshold is used to determined various characteristics. In the example of Figure 11, a wide QRS complex is used to identify a Ventricular Tachycardia, it will be appreciated that other features of an ECG signal may be used to identify Ventricular Tachycardia or other conditions. It is not required that the user of the method identifies a feature, e.g. a wide QRS complex, before applying the method; similarity between an input  
10 ECG signal and a reference signal known to relate to a cardiac condition is useable to identify that condition. To enable the method to be used for a wide range of situations, the threshold similarity is modifiable dependent upon the condition being considered.

Figure 12 shows the time evolution of the probability for an arrhythmic episode for the Random  
15 Forest classifier. The significance of the separation (between the 'arrhythmic' and 'normal' datasets in standard deviations) as a function of threshold probability is shown for the Random Forest classifier in Figure 13, which indicates that a threshold of 50% leads to a significance of roughly 1.9 standard deviations.

20 Figure 14 illustrates the distributions of abnormality fractions,  $F$  - for 'arrhythmic' and 'normal' patients for a Random Forest classifier. The distributions have been normalised to unit area for presentational purposes, where A.U. stands for arbitrary units.

An optimal decision boundary is arrived at by minimising the root mean square error, denoted  
25 as RMSE, and defined as:

$$RMSE = \sum_i (F_i - F_{decision})^2$$

(Equation 1.1)

Where  $F_i$  is the fraction of abnormal heartbeats for the  $i$ 'th misclassified patient and  $F_{decision}$  is the abnormality fraction under consideration.  $RMSE$  can be thought of as a measure of  
30 distance from the decision boundary for misclassifications.

A hybrid classifier may be created by combining the abnormality fractions,  $F$ , for each model listed above. The combination is a weighted sum defined as:

$$F_{hybrid} = \sum_j (w_j F_j)$$

(Equation 1.2)



Where  $w_j$  is the weight attributed to the  $j$ 'th classifier and  $F_j$  is the corresponding abnormality fraction,  $F$ . The weights,  $w_j$ , are determined according to the performance of the classifiers, as measured by their *RMSE* value.

5

More specifically, the weights,  $w_j$ , are determined dependent upon their *RMSE* value over misclassifications. The motivation for doing so is to achieve optimal performance of the resulting hybrid classifier in an unbiased way. Other commonly used metrics could lead to the wrong weights being attributed to classifiers and, consequently, suboptimal decisions.

10

The performance of the method described herein may be determined according to a number of performance metrics, exemplary metrics are listed below:

- Accuracy (*A*), defined as:

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$

(Equation 1.3)

15

Where the numerator is a sum of true positives (*TP*) and true negatives (*TN*) and the denominator includes false positives (*FP*) and false negatives (*FN*).

- Sensitivity (*SE*), defined as:

$$SE = \frac{TP}{TP + FN}$$

(Equation 1.4)

20

- Specificity (*SP*), defined as:

$$SP = \frac{TN}{TN + FP}$$

(Equation 1.5)

- Precision (*P*), defined as:

$$P = \frac{TP}{TP + FP}$$

(Equation 1.6)

25

- The area under the receiver operating characteristic (ROC) curve.

The receiver operating characteristic (ROC) curve plots the true positive rate (sensitivity) as a function of the false positive rate (100 – specificity). The area under the ROC curve (AUROC) is useable as a performance metric, with a higher area relating to higher accuracy.

30

Figures 15 illustrate the performance of a “committee voter” classifier as has been described above, and a boosted decision tree voter, considering each of the above performance metrics. Each metric is shown against a threshold probability of a cardiac event occurring that relates to the dataset used.

More specifically: Figure 15a illustrates the accuracy; Figure 15b illustrates the sensitivity, Figure 15c illustrates the specificity; Figure 15d illustrates the ROC curve; Figure 15e indicates the precision; and Figure 15f illustrates the RMSE. It can be seen that the boosted decision tree has a working point with a higher sensitivity than the committee voter; however, the committee voter display better specificity and has a higher area under the ROC curve.

The decision of which classifier to use may depend upon the dataset, where the boosted decision tree may be used as long as the boosted decision tree is not operating close to the decision boundary associated with the boosted decision tree, and the committee voter is used if the boosted decision tree is operating close to the associated decision boundary.

The method described herein may be integrated with and/or implemented by existing patient monitoring equipment.

20

### System architecture

Figure 16 illustrates an example of a system for predicting cardiac events. A physiological data source 20 (e.g. means for providing physiological data), extracted from the cardiorespiratory system of a patient 22, is communicated to an analysis module 24, which analyses the extracted physiological data. This communication may occur over a wired connection or a wireless network. The physiological data source 20 can be, for example, an electrocardiogram (ECG) machine; a pulsometer; a wearable cardioverter defibrillator; an implantable cardioverter defibrillator; a respiratory monitor; and/or a capnography monitor.

30

The analysis module 24 is configured to evaluate the extracted physiological data, for example evaluating a property of multiple heartbeats in the data, and determine whether said property exceeds an abnormality threshold. This information is then used to derive a probability of the patient experiencing a cardiac event, for example using the method described above in relation to Figures 2a and 2b to evaluate said property and derive said property.

35

The analysis module 24 comprises a hybrid classifier trained and operating as described above in relation to Figure 2b. The module 24 may comprise part of a dedicated machine, for example



running locally to the patient and data source, or be part of a network, running on a server or in the “cloud”.

5 If the analysis module 24 determines that the probability of the patient experiencing a cardiac event in a subsequent time period is above some pre-defined threshold, then the analysis module will trigger a means for providing an output 26, for example an alarm, or other alert, that can alert a healthcare provider that the patient is at risk. This can enable the healthcare provider to take preventative action.

10 In some embodiments, the output comprises alerting a healthcare provider and providing medical records, such as ECG readings, to the provider. The provider may then be inclined to analyse the data further and/or maintain a close watch on the patient, so that rapid action can be taken if the patient does suffer a cardiac arrest. This output may comprise a ‘risk window’ in which a heightened risk is identified; a close watch could then be maintained in this period.

15

#### Display of output

The output displays one or more probabilities, as determined using the methods described. The probabilities are output in numerous forms, notably:

20 • A binary assessment is used as a threshold indicator, where a critical value triggers an alarm. This is particularly useful as a first indicator that a patient may require attention. A threshold here is used to indicate that urgent help is required, or that patient data should be looked at more closely. There may be multiple thresholds which each have a differing level of urgency.

25 • A probability of a cardiac event is output, where this allows a user to allocate resources, and make other decisions, appropriately. An uncertainty estimate is output alongside this probability.

30 In differing embodiments, the probability is output quantitatively (for example as a percentage risk) and/or qualitatively, (for example a patient may be categorised as one of low risk, medium risk, or high risk, where these correspond to probability ranges). A qualitative measure may be used to simplify the immediate interpretation by a user.

• A probability density function relating to the probability of an upcoming cardiac event is output, where this allows a user to more fully assess a situation.

35 These probabilities are typically used in conjunction so that, upon a threshold risk being passed, a user is directed to view a probability, or a probability function, to determine an appropriate action. This can then be used as a general indicator of a patient’s health, where an

increased likelihood of a cardiac event indicates that a patient is more likely to need attention during a certain period.

5 An uncertainty also being displayed further aids the determination of an appropriate action. A potential problem with any data based analysis, particularly an analysis of a complex situations, such as the prediction of a cardiac event, is that a precise result is rarely achievable; this leads to a figure (such as a probability) on its own having limited use – especially due to the difficulty in determining if this figure is reasonable. The inclusion of an uncertainty based measure (such as a variance, or error bounds), enables a better judgement to be made regarding any given figure/probability.

10 Advantageously, a probability enables a user to make a rapid assessment, as a probability is intuitively interpreted more easily than, for example, a risk score. Additionally, a probability density function gives a user a large amount of information in a concise format.

15 In various embodiments, probabilities are also output for a number of timeframes. An initial output is simply a probability without any time reference. A more useful output is a probable time-to-cardiac-event. More specifically, probabilities may be output for time ranges, where this allows efficient allocation of resources.

20 The outputting of probability density functions for numerous timeframes enables limited resources to be scheduled effectively: e.g. a limited number of staff can be directed to be ready to assist certain patients at times of increased risk; a probability density function may be used to assess whether a cardiac event is almost certain or whether the risk is more unpredictable.

25 In some embodiments, a probability density function is displayed numerically, where a mean, a standard deviation, and a kurtosis (indicating the skew of the distribution) are displayed. In these, or other, embodiments, the function is (also) displayed graphically.

30 There are, in some embodiments, numerous, user selectable, ways to illustrate a probability, for example a best fit normal distribution, a skew normal distribution, or a Poisson distribution. A preferred distribution is suggested during analysis, where a suitable distribution depends on, for example, the amount of information available.

35 In some embodiments, the probability assessment is continuously updated, where this occurs as relevant information is obtained. An initial assessment uses historic data, and/or admissions data; this initial assessment is then updated (and improved) using recorded and evaluated data (such as the RR intervals above) as it becomes available.



In preferred embodiments, a Bayesian probabilistic framework is used in this updating, where Bayesian inference is used to obtain a probability. This is related to a form of Bayes rule, which is displayed in equation 2.1 below:

5

$$P(Y|X, \alpha) = \frac{P(X|Y)P(Y|\alpha)}{P(X|\alpha)} \propto P(X|Y)P(Y|\alpha)$$

(Equation 2.1)

where:  $P(Y|\alpha)$  is the prior distribution (e.g. the previously calculated probability);

$P(Y|X, \alpha)$  is the posterior distribution ( e.g. the updated probability);

10  $P(X|\alpha)$  is the marginal likelihood ( e.g. the likelihood of the recently sampled data given the entire set of data);

$P(X|Y)$  is the sampling distribution ( e.g. the probability of the observed data given the current distribution); and

$\alpha$  is the statistical hyperparameter of the parameter distribution (e.g.  $Y \sim P(Y|\alpha)$ ).

15

This equation is used to derive an updated probability based upon a prior probability and the probability of the occurrence of the recently sampled data. Using this equation, recent data which is indicative of a cardiac event being likely would be more concerning in a patient previously judged to be high-risk than it would in a patient previously judged to be low-risk (an interpretation of this is that in the low-risk patient this data is more likely to be anomalous). The use of Bayesian inference is then useful for reducing the rate of false positives, as the prior probability will be small for low-risk patients.

20

Notably, in the given example, the occurrence of data indicative of a cardiac event would be unlikely given the prior distribution, and so this would have a significant effect on the posterior distribution. Due to this, the data would not simply be written off entirely as anomalous; while it may not immediately result in a warning, continued occurrence of data indicative of a likely cardiac event would rapidly increase the probability (so that the chance of missing a cardiac event is unlikely); however, advantageously, a single (potentially anomalous) datapoint would not trigger a false positive warning.

25

30

To further reduce the likelihood of false negatives, in some embodiments, a Bayesian inference model is used alongside a threshold marginal likelihood: a marginal likelihood which is indicative of a very high chance of an upcoming cardiac event then triggers a warning even if the overall probability remains low due to a consistently low prior probability.

35

The updating of the probability takes place periodically (for example each five seconds, or each minute), where a longer update (or refresh) period use less computing power. This update period is, in some embodiments, small enough that the probability is updated effectively continuously (i.e. the period is so small as to not be noticeable by a user).

5

In some embodiments, there is a component within the apparatus which allows a choice of the update period – this may also be selectively determined based on the use of the apparatus (where an implanted device may prioritise battery longevity over rapid updates).

10 A consideration here is that, in many situations, it is possible to maintain an accurate probability while making only periodic updates, especially where there is a large prior distribution (i.e. where measurements have been taken for a long time). The update period is then based upon the prior distribution. As an upper limit for the time, these updates may be limited, so as to be regular enough that they do not miss a cardiac event.

15

Figure 17 shows a component diagram for analysing patient data and displaying an output.

20 One or more measurement device(s) (e.g. an ECG, a patient file) 32 transmit(s) data to a local server 34. These data are then transmitted to a network server 36, and fed through an analysis module 24 (as discussed, e.g. with reference to Figure 2). The output of the classifier passes through a results formatter 40 before being transmitted back to the local server 34 (this results formatter 40 format results to be output as a warning alarm, or a display of probability). The output is then presented on a UI 42 for one or more users, this uses, for example, a smartphone, a screen, or a display distributed by a hospital. In some embodiments, this also  
25 comprises a speaker, which provides an audible output if a threshold probability is exceeded.

30 By sending data via a network server 36, instead of storing all data on a local device, the data can be displayed to numerous users simultaneously. This allows the gathering multiple opinions, or to alert numerous users simultaneously, so that the user in the best position to may be notified.

The use of a network server 36 also enables remote monitoring of a patient. This may be used for a patient with an implantable device, where data recorded by the device is transferred to a network server 36, evaluated by the analysis module 24, and then displayed on a UI 42 to both  
35 the user and (separately) a healthcare professional, who may then check on the user at an appropriate time.



The figures as described above show a system for monitoring a patient. As a general overview: in Figure 16, there is a patient 22, for which it is desired to output a probability of a cardiac event. A means for providing physiological data 20, such as an electrocardiogram (ECG) is used to obtain this data. Typical data is shown in Figure 1; specific data, such as the RR intervals is extracted from this data. This data is then fed into an analysis module 24, which is discussed with reference to Figures 2b and 2c.

The analysis module 24 is provided with the specific data (the RR intervals) as in Figures 2b and 2c. Processing then occurs:

- 10 1. outliers are removed. This is demonstrated by Figure 3;
2. numerous properties are determined, such as the mean RR interval, and the ectopic beat frequency;
3. optimal context lengths are determined for each property (or for a grouping of properties, such as time domain properties). This is demonstrated by Figures 7a and 7b;
- 15 4. the optimal length of data is fed into an artificial intelligence based classifier;
5. the artificial intelligence based classifier, which is formed of multiple different classifiers combined to obtain a hybrid classifier, determines threshold abnormality values for each property, which are indicative of an upcoming cardiac event (e.g. a threshold mean RR interval is calculated, where a mean interval above this threshold is indicative of an upcoming cardiac event).

20 The threshold values are determined based upon past data from multiple sources, for example a database containing physiological data for patients alongside occurrences of cardiac events may be used for training a classifier.

25 6A. The data which has been fed in to the classifier is compared to the relevant threshold, and a probability of a cardiac event occurring is determined (based on the fraction of the data which exceeds the threshold). This probability is displayed, and an alarm is sounded if a high probability of a cardiac event is obtained.

30 6B. The data which has been fed in to the classifier is output to an optimisation stream, where it is used to further optimise the determination of following threshold values (i.e. it is incorporated into the training set).

This output is then presented using a means for providing an output 26.

The means for providing physiological data, and the means for providing an output are described in more detail above with reference to Figure 17.

35

Figure 18a shows an ECG reading for a patient suffering a sudden cardiac arrest. The episode begins at 427.5s and the patient is defibrillated roughly 1s later. Figure 18b shows the output of the system of Figure 16 when applied to the ECG reading of this patient prior to the event.

5 It can be seen from Figure 18b that this apparatus would have identified an increased risk of a cardiac event at 190s and thus could have been used to produce an alert around 4 minutes before the occurrence of the event at 429s. This alert could be used to indicate, for example, that a closer watch should be kept on the patient for a certain amount of time or that a medical practitioner should review the patient's data in detail.

10

### Alternatives and modifications

#### Data types

The use of RR intervals is an example of a type of data – more specifically a type of physiological data, yet more specifically a type of cardiac data – usable with the described methods; more generally, any type of patient data, or any combination of types of data could be used with these methods, where the use of a combination of patient data may lead to fewer false positives (or false negatives). Examples of preferred types of data are (with some overlap as, for example, telemetry records and clinical data both comprise physiological data):

- telemetry records, such as arterial blood pressure, pulse contour data, or pulse rate;
- 20 • demographic data, such as age, sex, or race (this may come from an electronic health report/patient profile);
- Admission/historic data, such as a recent illness or any history of illness; in particular concomitant conditions, such as emphysema or diabetes;
- clinical data, such as haemoglobin values;
- 25 • laboratory data, such as the results of tests;
- imaging data, such as x-rays or MRI scans.

Where multiple data types are considered, each of these types of data is treated similarly to the RR intervals: properties (such as a mean or a standard deviation) are extracted, and an optimal context length for these features determined – as an example, there is an optimal length of patient history to consider, where data more than, for example, 10 years old may have a negligible contribution to a prediction of future health. Numerous data types are considered in the determination of a probability, where, in some embodiments, each data type has a different weighting (where this weighting is based upon historic data and determined by the classifiers).

35

In various embodiments, the data types used are optimised, where this is used within the display of a probability. In each situation, there is selected a combination of data features with



the most significant effect; this is particularly useful where an implantable device is used, and using a low number of data types is desirable, as this minuses the computational burden.

5 In some embodiments, to avoid the need for new measuring equipment, analysis occurs only using data which is attainable using current measuring methods.

10 While the data recording methods discussed have primarily involved specialist equipment (e.g. electrocardiograms), the methods discussed could equally be used with other, more widely available equipment. As an example, there exist many user wearable devices which are used to monitor a heartrate or a pulse (such as a Fitbit™). The data recorded using this, or a similar, device could be used with the AI classifier described above to obtain a probability of a cardiac event, or to output a general health measure. If used in such a device, the output may be a displayed probability, or measure of health, to the user, or an automatic warning sent to, for example, an ambulance, if a threshold probability is exceeded. This may be particularly useful  
15 in devices such as a Fitbit™, which are used during periods of increased activity (where stress may be placed upon the heart).

#### Context length determination

20 The context length determination has been explained using the example of a  $\chi^2$ -test ('chi-squared' test); other tests could also be used for this determination. Various embodiments use one of (or a combination of): a Kolmogorov-Smirnov test, a comparison of the moments of distributions, or an Energy Test (as described by Guenter Zech and Berkan Aslan).

25 When using the Energy Test an Energy Test metric, T, is computed between two distinct unbinned multivariate distributions. One such example is arrhythmic and normal heartbeat distributions, which give a non-zero T-value. This is used in some embodiments as an additional test on the probability of a cardiac event: an Energy Test is performed and a T-value calculated, this T-value is updated after each heartbeat and a warning is issued if the T-value exceeds a predetermined threshold (which is based on past data, and may be determined for  
30 each patient based upon their specific data). The context length over which the Energy Test is performed is determined as with any other dataset. This test may be used in isolation, or in conjunction with any other method described, where use in conjunction with other methods may reduce the likelihood of false negatives or false positives.

#### 35 Autoregressive models

In some embodiments, autocorrelation is considered along with a measure of the lag required to obtain an autocorrelation. As an example, in the short term, the occurrence one cardiac

event may be indicative of another cardiac event being likely to occur (i.e. recent cardiac events may have high autocorrelation), as these events are often related to periods of otherwise poor health. In the long term, a previously occurring cardiac event (e.g. a cardiac event which occurred in a previous year), may be a poor indicator of a subsequent cardiac event (i.e. distant cardiac events may have low autocorrelation), as the period of poor health may have passed. The suitability of using an autoregressive model is determined by comparing these correlations and lags.

A consideration with autocorrelation is that (useful) autocorrelation may be negative or positive. In the previously used example, it may be the case that a previous, but distant cardiac event (e.g. one that occurred in a previous year), is a good indicator that a cardiac event is unlikely, as the person may have worked to improve their health in response to the previous event.

#### Other conditions

The methods described could be used for a range of other conditions, for example, as well as a cardiac event, indicators of an upcoming arrhythmia may also be used to predict a stroke. The methods disclosed herein could also be used to measure conditions away from the heart: the flow of blood could, for example, be monitored as relates to transfer to the brain. In this situation, a context length would still be of relevance: monitoring the blood flow into the brain could be used to give a prediction of brain related events (such as brain aneurysms).

More generally, the methods disclosed could be used as a general indication of health. Abnormal operation of any pulse based condition is a possible indicator of not only the probability of a specific event (e.g. arrhythmia), but also that the patient is likely to be at heightened risk of a more general health-related incident. These methods may then be used to indicate that a patient may need more careful monitoring during a determined period, or that it may be valuable to analyse patient data in more detail and/or to carry out tests.

It will be understood that the invention has been described above purely by way of example, and modifications of detail can be made within the scope of the invention.

Each feature disclosed in the description, and (where appropriate) the claims and drawings may be provided independently or in any appropriate combination.

Reference numerals appearing in the claims are by way of illustration only and shall have no limiting effect on the scope of the claims.



## Claims

1. A method of analysing cardiac data relating to a patient, comprising  
providing cardiac data relating to the patient;  
5 determining a property of the data, wherein the property is determined over a particular context length, the context length being selected based on the property;  
comparing one or more features of the property against a predetermined threshold value, thereby to indicate a probability of the patient experiencing a cardiac event; and  
providing an output based on the comparison.  
10
2. The method of Claim 1, further comprising modelling the property using a function; and wherein comparing the one or more features of the property against the predetermined threshold value comprises comparing one or more descriptors of the function against a predetermined descriptor threshold value.  
15
3. The method of Claim 2, wherein determining a property of the data comprises determining a plurality of datapoints related to the property; and wherein modelling the property using a function comprises modelling the distribution of the datapoints using a function.
- 20 4. The method of Claim 2 or 3, wherein modelling the property using a function comprises determining a probability density function suitable for modelling the property.
- 25 5. The method of any of Claims 2 to 4, wherein modelling the property using a function comprises superposing one or more Gaussian functions, preferably superposing Gaussian functions of equal surface.
6. The method of any of Claims 2 to 5, wherein comparing one or more descriptors of the function comprises comparing at least one of: a mean; a variance; and a kurtosis.
- 30 7. The method of any preceding claim, further comprising providing contextual data relating to the patient; wherein the threshold value is dependent upon the contextual data.
8. The method of any preceding claim, further comprising:  
comparing a further property against a predetermined contextual threshold value,  
35 wherein the contextual threshold value is dependent upon contextual data; and  
providing an output based on both the comparison of the property and the comparison of the further property.

9. The method of Claim 7 or 8, wherein the contextual data comprises at least one of: historic data related to the patient, an electronic health record related to the patient, physical characteristics of the patient; and demographic characteristics of the patient.
- 5 10. The method of any preceding claim, comprising  
representing the data as a series of fixed size representations;  
providing an attention mechanism arranged to identify one or more points of interest  
within the data based upon the fixed size representations; and  
providing an output based on the identified points of interest.
- 10 11. The method of Claim 10, wherein representing the data as a series of fixed size  
representations comprises using a network operating over fixed-sized windows of data,  
preferably wherein the network is a neural network and/or a long short-term memory  
network.
- 15 12. The method of any preceding claim, wherein the threshold value is determined based on a  
dataset comprising a plurality of data obtained from multiple sources.
- 20 13. The method of any preceding claim, wherein the or each property is determined over a  
context length which is an optimally discriminating context length for that property.
- 25 14. The method of any preceding claim, wherein the properties comprise at least one of: a  
mean; a standard deviation; a standard deviation in successive differences; a measured  
heart rate variability (HRV) of a patient; and a fraction of multiple heartbeats that exceed  
an abnormality threshold.
- 30 15. The method of any preceding claim, wherein the predetermined threshold is determined by:  
training at least two classifiers to classify a property of multiple heartbeats within the  
cardiac data using at least one machine learning algorithm; and  
combining the at least two classifiers to produce a hybrid classifier;  
wherein the combination is based on a performance metric.
- 35 16. A method of training a hybrid classifier for analysing cardiac data related to a patient, the  
method comprising the steps of:  
training at least two classifiers to classify a property of multiple heartbeats within the  
cardiac data using two or more different machine learning algorithms; and  
combining the at least two classifiers to produce a hybrid classifier;  
wherein the combination is based on a performance metric.



17. The method of Claim 15 or 16, further comprising  
determining a best performing classifier and a second best performing classifier based upon a performance metric;  
5 outputting the classification of the best performing classifier when the output of the best performing classifier is not close to a decision boundary; and  
outputting the classification of the second best performing classifier when the output of the best performing classifier is close to the decision boundary;  
optionally wherein the output of the best performing classifier is considered to be not  
10 close to the decision boundary when a threshold probability of a correct classification is exceeded.
18. The method of any of Claims 15 to 17, wherein training at least two classifiers comprises combining at least two trained classifiers to produce a hybrid classifier;  
15 wherein combining the at least two trained classifiers comprises applying weightings to each classifier based on a performance metric associated with each respective classifier.
19. The method of any of Claims 15 to 18, wherein the performance metric comprises at least one of: an accuracy; a sensitivity; a specificity; a precision; and an area under a receiver  
20 operating characteristic (ROC) curve.
20. The method of any of Claims 15 to 19, wherein training at least two classifiers comprises using a genetic algorithm and/or simulated annealing.
- 25 21. The method of any of Claims 15 to 20, wherein training a classifier comprises:  
providing annotated cardiac data, wherein the annotation indicates the occurrence of one or more cardiac events;  
training a detection classifier to detect cardiac events using the annotated cardiac data;  
labelling unannotated cardiac data using the trained detection classifier;  
30 training a classifier to classify a property of multiple heartbeats using the labelled cardiac data;  
optionally wherein labelling unannotated cardiac data using the trained detection classifier comprises labelling a subset of unannotated cardiac data dependent upon a threshold probability of correctness.
- 35 22. The method of any of Claims 15 to 21, further comprising:  
providing a reference dataset of annotated cardiac data;  
providing an input dataset of unannotated cardiac data;

normalising each member of the reference dataset and each member of the input dataset to have the same dimensions;

comparing each normalised member of the input dataset with one or more normalised members of the reference dataset to identify a measure of similarity;

5 determining labels for the input dataset dependent upon the respective measures of similarity;

training a classifier to classify a property of multiple heartbeats using the labelled cardiac data;

10 optionally wherein comparing each normalised member of the input dataset with one or more normalised members of the reference dataset comprises determining a root mean square error (RMSE).

23. The method of any of Claims 15 to 22, wherein the cardiac data comprises ECG signals.

15 24. A system for analysing cardiac data relating to a patient, comprising:

means for providing cardiac data relating to the patient;

an analysis module for determining a property of the data, wherein the property is determined over a particular context length, the context length being selected based on the property;

20 a comparison module for comparing the property against a predetermined threshold value, thereby to indicate a probability of the patient experiencing a cardiac event; and  
a presentation module for providing an output based on the comparison.

25 25. The system of Claim 24, wherein the analysis module comprises a hybrid classifier trained according to the method of any of Claims 15 to 23.





**Application No:** GB1814620.9

**Examiner:** Peter Burns

**Claims searched:** 1-15 & 24; 25 in part

**Date of search:** 28 February 2019

**Patents Act 1977: Search Report under Section 17**

**Documents considered to be relevant:**

Category	Relevant to claims	Identity of document and passage or figure of particular relevance
X,E	1-4, 6-9, 12-15 & 24-25	GB 2560339 A (SIRENDI et al.) whole document is relevant, see e.g. pages 2 & 4 (final paragraphs) in particular
X	1-14 & 24	US 2005/0137482 A1 (LAITIO et al.) see in particular paragraphs 0010-0013 & 0031-0033
X	1-3, 7-13 & 24	US 2017/0196769 A1 (FREEMAN et al.) see paragraphs 0022 & 0088 in particular
X	1-3, 7-13 & 24	WO 2016/183683 A1 (CHAUHAN & SUSZKO) see in particular paragraphs 0093, 0098 and claims 1 & 7-8 of the document
X	1-3, 7-13 & 24	US 2002/0068875 A1 (SCHROEPEL et al.) refer particularly to paragraphs 0006, 0009, 0011, 0043-0045 & 0062
X	1-3, 7-13 & 24	EP 3120753 A1 (NOGAMI & TAKIZAWA) refer e.g. to claims 1-2 of the document, and paragraph 0002

**Categories:**

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

**Field of Search:**

Search of GB, EP, WO & US patent documents classified in the following areas of the UKC<sup>X</sup> :

--

Worldwide search of patent documents classified in the following areas of the IPC

A61B
------

The following online and other databases have been used in the preparation of this search report

EPODOC, WPI
-------------



Intellectual  
Property  
Office

**International Classification:**

<b>Subclass</b>	<b>Subgroup</b>	<b>Valid From</b>
A61B	0005/00	01/01/2006
G06K	0009/62	01/01/2006
G16H	0050/20	01/01/2018