

Features characterising cardiac autonomic neuropathy in diabetes using ensembled classification

Nedergaard, Rasmus Bach; Scott, Matthew; Wegeberg, Anne Marie; Okdahl, Tina; Størling, Joachim; Brock, Birgitte; Drewes, Asbjørn Mohr; Brock, Christina

Published in:
Clinical Neurophysiology

DOI (link to publication from Publisher):
[10.1016/j.clinph.2023.06.011](https://doi.org/10.1016/j.clinph.2023.06.011)

Creative Commons License
CC BY 4.0

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Nedergaard, R. B., Scott, M., Wegeberg, A. M., Okdahl, T., Størling, J., Brock, B., Drewes, A. M., & Brock, C. (2023). Features characterising cardiac autonomic neuropathy in diabetes using ensembled classification. *Clinical Neurophysiology*, 154, 200-208. <https://doi.org/10.1016/j.clinph.2023.06.011>

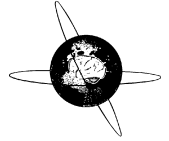
General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.



Features characterising cardiac autonomic neuropathy in diabetes using ensembled classification



Rasmus Bach Nedergaard^{a,*}, Matthew Scott^a, Anne-Marie Wegeberg^{a,b}, Tina Okdahl^a, Joachim Størling^c, Birgitte Brock^c, Asbjørn Mohr Drewes^{a,b,d,e}, Christina Brock^{a,d,e}

^a Mech-Sense, Department of Gastroenterology and Hepatology, Aalborg University Hospital, Aalborg, Denmark

^b Thisted Research Unit, Aalborg University Hospital Thisted, Thisted, Denmark

^c Department of Clinical Research, Steno Diabetes Center Copenhagen, Gentofte, Denmark

^d Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

^e Steno Diabetes Center Nordjylland, Aalborg, Denmark

See Editorial, pages 198–199

HIGHLIGHTS

- It is possible to obtain good classification results of cardiovascular autonomic neuropathy using supervised machine learning.
- The best classifier is the random forest classifier with a F1-score of 60% on the reduced dataset.
- In the ensembled data no standard clinical biochemical analytes were present in the 10 best classifiers.

ARTICLE INFO

Article history:

Accepted 3 June 2023

Available online 30 June 2023

Keywords:

Diabetes
Machine learning
Support vector machine
Decision tree
Random forest
Ensembling

ABSTRACT

Objective: Using supervised machine learning to classify the severity of cardiovascular autonomic neuropathy (CAN). The aims were 1) to investigate which features contribute to characterising CAN 2) to generate an ensembled set of features that best describes the variation in CAN classification.

Methods: Eighty-two features from demographic, beat-to-beat, biochemical, and inflammation were obtained from 204 people with diabetes and used in three machine-learning-classifiers, these are: support vector machine, decision tree, and random forest. All data were ensembled using a weighted mean of the features from each classifier.

Results: The 10 most important features derived from the domains: Beat-to-beat, inflammation markers, disease-duration, and age.

Conclusions: Beat-to-beat measures associate with CAN as diagnosis is mainly based on cardiac reflex responses, disease-duration and age are also related to CAN development throughout disease progression. The inflammation markers may reflect the underlying disease process, and therefore, new treatment modalities targeting systemic low-grade inflammation should potentially be tested to prevent the development of CAN.

Significance: Cardiac reflex responses should be monitored closely to diagnose and classify severity levels of CAN accurately. Standard clinical biochemical analytes, such as glycaemic level, lipidic level, or kidney function were not included in the ten most important features. Beat-to-beat measures accounted for approximately 60% of the features in the ensembled data.

© 2023 International Federation of Clinical Neurophysiology. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

It is estimated that diabetes affects more than 535 million people globally, with projections suggesting that upwards of

783 million people will be affected in 2045 (Sun et al., 2022). As a result, clinical relevant micro- and macro-vascular complications are also expected to increase (Sinnreich et al., 2005). Among the lesser understood microvascular complications, is diabetic

* Corresponding author.

E-mail addresses: r.nedergaard@rn.dk (R.B. Nedergaard), matthew.scott18@imperial.ac.uk (M. Scott), a.wegeberg@rn.dk (A.-M. Wegeberg), t.okdahl@rn.dk (T. Okdahl), joachim.stoerling@regionh.dk (J. Størling), birgitte.brock@regionh.dk (B. Brock), amd@rn.dk (A.M. Drewes), christina.brock@rn.dk (C. Brock).

<https://doi.org/10.1016/j.clinph.2023.06.011>

1388-2457/© 2023 International Federation of Clinical Neurophysiology. Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

autonomic neuropathy, which among others affects the cardiovascular systems (Vinik et al., 2013). Though autonomic neuropathy of the cardiovascular system (CAN) increases the risk of silent myocardial ischemia and mortality (Feldman et al., 2019; Pop-Busui, 2012), it is still not clearly understood why some develop this, while others do not.

Precision diagnostics tailor clinical diagnosis to the individual by use of biomedical information coupled with machine learning. This approach can facilitate earlier diagnosis and individual risk (Chung et al., 2020). Clinically, this has been used to gain insights into diabetic sensorimotor polyneuropathy (Haque et al., 2022), general complications (Ljubic et al., 2020), and recently CAN (Abdalrada et al., 2022).

The use of machine learning allows for the selection of features from large samples of complementary clinical data. To improve generalisability and minimise overfitting to a particular training set, feature selection techniques must be employed. Support vector machines (SVMs) are well-established algorithms for performing classification on both linear and non-linear data, often in clinical settings (Kumari and Chitra, 2013; Othman, 2011). A common issue involving SVMs is their operation as a 'black box' (Cherkassky and Dhar, 2015). For this reason, medical analysis is often performed with hierarchical models (Burgess et al., 2000; Wang and Gatsonis, 2008) to eliminate the uncertainty in the weighting of variables that produce a classification. A common hierarchical model (in this case, a hierarchical clustering model) is a decision tree, which has merited medical applications in the diagnosis of type 2 diabetes (Habibi et al., 2015; Al Jarullah, 2011). This model classifies data by producing a condition on individual features or nodes to produce two sub-nodes that stem from each decision. By following this tree from the topmost node and following the numerical conditions on each node until an end node is met, a prediction can be made for each feature set. In reducing the issue of overfitting with decision trees, the technique of ensemble learning is often applied. Ensembling such trees produce a random forest (Breiman, 2001), which has similarly been applied to the risk prediction of type 2 diabetes (Xu et al., 2017). With the introduction of multiple models, a metric for aggregating the importance's over each model can be utilised to preserve the features that are consistently useful in modelling. Such approaches are referred to as ensembling, much like the intuition of random forests, and have seen both theoretical (Seijo-Pardo et al., 2015) and commercial use (Shah and Peretiakto, 2021).

We hypothesised that the severity of CAN may be classified based on clinical features, heart rate variability measures, and biochemical measures including inflammatory markers using a combination of supervised machine learning techniques and aggregation of the best performing features.

The aims were to 1) investigate which clinical and paraclinical features can help explain the characteristics of CAN, and 2) generate an ensemble set of features that best describes the variation in CAN severity.

2. Method

2.1. Study population

This cross-sectional dataset was comprised from two cohorts of people with diabetes recruited from the Department of Endocrinology, Aalborg University Hospital: Cohort 1: Cross-sectional study investigating cardiovascular and autonomic complications in consecutive patients attending the annual health visit with unknown status of diabetic neuropathy (type 1 diabetes, $n = 56$; Type 2 diabetes, $n = 100$) and Cohort 2: Baseline data from a randomized

controlled trial (RCT) (Brock et al., 2019) in 48 participants with type 1 diabetes and verified peripheral neuropathy according to the Toronto criteria. The studies comprising the current dataset were conducted in accordance with the Declaration of Helsinki, local regulations, and International Conference on Harmonization Good Clinical Practice guidelines. The North Denmark Region Committee on Health Research Ethics, Denmark approved both studies (N-20170045 and N-20130077). Furthermore, Cohort 2 was approved by the Danish Health and Medicines Authority (EudraCT No.: 2013-004375-12). All participants gave their written informed consent before participating in any data recording. The clinical characteristics of the cohorts is available in Table 1.

2.2. Clinical measures

Characteristics of the dataset consist of age, sex, disease duration, and smoking habits. Height was measured using a stadiometer (Seca GmbH & Co. KG., Hamburg, Deutschland). Weight was assessed with a standardised calibrated scale (BWB-800AS, Tanita, Arlington Heights, Illinois, USA). Body mass index was calculated based on height and weight. Cardiac measures (Heart rate, resting systolic, and diastolic blood pressures) were obtained in a seated position using a blood pressure monitor (Intellisense®, Omron Healthcare, Inc., Bannockburn, Illinois, USA). Additionally, a vibration perception threshold of the first toe was performed on the side of the dominant hand.

2.3. Clinical biochemistry

On the day of examination fasting blood samples for assessment of: blood glucose, haemoglobin A1c, lipid profiles (cholesterol, LDL cholesterol, HDL cholesterol, and triglycerides), haematology (erythrocytes, haemoglobin, thrombocytes, leucocytes, iron, transferrin, transferrin saturation, ferritin, cobalamin), renal function (estimated glomerular filtration rate, creatinine, sodium, albumin, potassium, and uric acid), liver function (alanine transaminase, alkaline phosphatase, bilirubin), pancreatic function (amylase), and thyroid function (thyroid-stimulating hormone, thyroxine T4, triiodothyronine T3). The blood samples were obtained according to the standard laboratory procedures and analysed at the Department of Clinical Biochemistry, Aalborg University Hospital.

2.4. Assessment of cardiovascular autonomic neuropathy

All participants underwent cardiovascular autonomic reflex testing in the morning (between 8:00 and 10:00 a.m.) in a consistently lit, quiet laboratory. This was done to minimize variance in circadian rhythm. The diagnosis of CAN was performed using either orthostatic hypotension measure or The Vagus™ device (Medicus Engineering ApS, Aarhus, Denmark). Validation of the Vagus™ has been described elsewhere (Ejskjaer et al., 2008; Fleischer et al., 2011). Following 10 minutes of rest a CAN score was calculated based on three recordings using the Vagus™ device: electrocardiographic recordings at rest, and during cardiovascular autonomic reflex tests consisting of deep breathing (expiration/inspiration ratio), The valsalva ratio, and postural change (30:15 supine to standing ratio). To help mitigate missing data if one of the tests were not completed, the CAN score was estimated from the remaining two procedures. The state established CAN (CAN 2) was defined if two or more abnormal reflex tests were registered, or orthostatic hypotension defined as > 20 mmHg reduction in systolic blood pressure on standing; borderline CAN (CAN 1) was selected if one abnormal reflex test was present, and no CAN

Table 1

Demographic and variables of the study cohorts, divided over the three cardiovascular autonomic neuropathy (CAN) grades 0, 1, and 2. A one-way ANOVA using CAN as a categorical independent variable to test differences of variables, gender and diabetes type was tested using Pearson's Chi-squared test. For a description of cardiovascular autonomic neuropathy (CAN) see: "Assessment of cardiovascular autonomic neuropathy" For a description of biochemical measures see: "Clinical biochemistry" For a description of heart rate variability see: "Heart rate variability". Body mass index (BMI), alanine aminotransferase (ALAT), low-density lipoprotein (LDL), high-density lipoprotein (HDL), estimated glomerular filtration rate (eGFR), Root mean square of successive differences (rMSSD), standard deviation of the averages of RR (SDANN), standard deviation of the averages of RR (SDNN), mean standard deviation of the averages of RR for each 5-minute interval (SDNNi), very low frequency (VLF), low frequency (LF), high frequency (HF), The inflammation markers are tumour necrosis factor (TNF) and IL- are serum concentrations of cytokines (interleukin).

CAN; Fraction of dataset	0; 45.1%	1; 22.55%	2; 32.35%	Difference
Demographics				
Gender (M/F)	50/42	30/16	47/19	p = 0.087
Diabetes type (T1/T2)	34/58	15/31	55/11	p = <0.001
Age (years)	55 ± 17.6	57 ± 13	52 ± 11.3	p = 0.197
Disease duration (years)	15 ± 11.6	16 ± 11.5	28 ± 12.3	p = <0.001
Height (cm)	171 ± 8.9	173 ± 8.9	177 ± 9.3	p = <0.001
Weight (kg)	86.2 ± 18.4	88.7 ± 14.5	88.6 ± 17.4	p = 0.598
BMI (kg/m ²)	29.5 ± 5.9	29.8 ± 5.4	28.2 ± 4.7	p = 0.247
HbA1c (mmol/mol)	56.4 ± 9.7	61.6 ± 10.7	64.5 ± 13.9	p = <0.001
Biochemistry				
Creatinine (μmol/L)	75.8 ± 27.9	76.8 ± 19	83.1 ± 27.4	p = 0.198
Cholesterol (mmol/L)	4 ± 0.9	4.2 ± 0.9	4.4 ± 0.8	p = 0.052
Triglycerides (mmol/L)	1.4 ± 1	1.6 ± 1.2	1.2 ± 1	p = 0.122
ALAT (U/L)	28.8 ± 16.9	27.9 ± 15.8	26 ± 13.7	p = 0.54
HDL (mmol/L)	1.4 ± 0.4	1.4 ± 0.5	1.6 ± 0.5	p = 0.129
LDL (mmol/L)	2.1 ± 0.8	2.1 ± 0.7	2.3 ± 0.6	p = 0.054
eGFR (ml/min)	82.9 ± 12.9	82.5 ± 12.4	79.4 ± 14.4	p = 0.249
Heart				
rMSSD	37.4 ± 17.2	27.2 ± 7.8	21.8 ± 10.1	p = <0.001
SDANN	119.7 ± 41.4	103.3 ± 26.2	103.8 ± 29.5	p = 0.005
SDNN	141.3 ± 42.1	115.7 ± 31.1	114.4 ± 32.3	p = <0.001
SDNNi	54.2 ± 18.7	45.5 ± 13.2	40.9 ± 16.7	p = <0.001
VLF	1726.7 ± 1134.7	1202.2 ± 723.7	1116.9 ± 669.7	p = <0.001
LF	949.1 ± 818.2	533.2 ± 433	458.6 ± 363.8	p = <0.001
HF	474 ± 474.7	217.7 ± 148.6	173.9 ± 166.3	p = <0.001
Inflammation (pg/mL)				
TNF-α	1.5 ± 1.2	1.4 ± 0.5	1.5 ± 0.6	p = 0.759
IL-6	1.3 ± 3.2	1.5 ± 4.1	1 ± 0.9	p = 0.665
IL-8	13.9 ± 11.1	13.2 ± 5.2	13 ± 5	p = 0.773
IL-10	0.4 ± 0.5	0.3 ± 0.1	0.4 ± 0.5	p = 0.414

(CAN 0) was used when tests were within the normal range of the specific age-dependent cut-off values (Spallone et al., 2011).

2.5. Heart rate variability

Holter monitoring was performed using either the ePatch electrocardiographic Recorder (BioTelemetry Technology ApS, Hørsholm, Denmark) or Holter monitor (Lifecard CF; Del Mar Reynolds, Spacelabs Healthcare, Snoqualmie, WA, USA). This monitoring was performed for 24 hours. The monitor was mounted in the morning, and heart rate variability parameters were computed for the entire 24 hour recording period. Data was analysed using Cardioscope™ (SmartMedical, Gloucestershire, United Kingdom) for the ePatch, or Pathfinder (Software revision B code; Spacelabs Healthcare, Hertford UK) for the Lifecard CF system. The time-derived measures of HRV were: Standard deviation of all normal RR (SDNN), standard deviation of the averages of RR (SDANN), mean standard deviation of the averages of RR for each 5-minute interval (SDNNi), and root mean square of the successive differences (RMSSD). The frequency domain measures of HRV were: Very low frequency (VLF), low frequency (LF), and high frequency (HF). The frequency measures were analysed and interpreted according to (Malik et al., 1996).

Additionally, cardiac vagal tone (CVT), was analysed using the ProBioMetrics online app version 1.0 (ProBioMetrics, Kent, UK). Artifacts in the recordings were defined as a change in two succeeding QRS complexes exceeding 15 beats per minute, files were inspected and cleaned by removing five heartbeats before and after

to derive and artefact free CVT. The data was removed from analysis if the number of removed heartbeats exceeded 20%.

2.6. Inflammatory markers

On the day of examination blood samples were drawn from the cubital vein of fasting participants. The blood was centrifuged at 2500 rpm for 10 minutes, the serum was stored in a −80 °C freezer. Biomarker concentrations in plasma samples were analysed using the V-PLEX Neuroinflammation Panel 1 Human Kit (Meso Scale Diagnostics® [MSD], Gaithersburg, MD, USA) on a MESO QuickPlex SQ 120 instrument (MSD) according to the manufacturer's specifications. The analysis was conducted at the CoreLab of Steno Diabetes Center Copenhagen. Serum concentrations of cytokines (interferon-γ, tumor necrosis factor (TNF)-α, TNF-β, eotaxin, eotaxin-3, IFN-γ-induced protein (IP)-10, interleukin- (IL-) 1α, −1β, −2, −4, −5, −6, −7, −8, −10, −12/−23p40, −13, −15, −16, −17A, macrophage-derived chemokine (MDC), monocyte chemoattractant protein (MCP)-1, MCP-4, macrophage inflammatory protein (MIP)-1α, MIP-1β, CC chemokine ligand 17 (CCL17), and CRP. Sample values below the detection limit of the assay were assigned a value of the detection limit divided by $\sqrt{2}$. If more than 30% of the measured samples for any given biomarker were below the detection limit, the biomarker was excluded from the analysis. Likewise, samples with a coefficient of variation > 30% between duplicate measurements were excluded from the analysis (Croghan and Egeghy, 2003).

2.7. Pre-processing

Any missing fields were filled with the mean value from the feature of the respective diabetic disease (type 1 or type 2). After pre-processing missing values by averaging the remaining data over diabetes type two datasets are yielded: 1) The full dataset, which comprises all demographic, heart, blood sample, and inflammation data ($n = 82$) and 2) the reduced dataset, which only consists of the most important features produced by the feature extraction techniques ($n = 10$), this is visualised in Fig. 1. Each feature selection technique hence generates a separate reduced dataset that is independently analysed.

The use of feature selection in dataset reduction has gained success when the dataset is reduced to a minimum size of \sqrt{N} (Hastie et al., 2009) or $\log_2(N + 1)$ (Breiman, 2001) for use in random forest classifiers. This size may require an increase to $3 \log_2(N + 1)$ with many categorical (discrete) variables. However, with a dimensionality of 82 features, Breiman 2021 (Breiman, 2001) suggests that $\sqrt{82} \approx 9$, or $\log_2(82 + 1) \approx 6$ features are required for a suitable model. A total of 10 features are thus offered to capture greater dataset variance and offer more potential for the exploration of the features that may influence CAN Score.

2.8. Data analysis

All analysis was performed using the Python3 scikit-learn (Pedregosa et al., 2011) library and all figures were produced using the “matplotlib” (Hunter, 2007) library. A variety of classification techniques documented below are investigated using the full dataset with a 20% test-train split, with all results (accuracies, features,

and confusion matrices) averaged over 1000 iterations. The key features that contribute to each classifier are then documented and aggregated to establish which clinical trials are most important in classifying the CAN scores of a patient.

2.8.1. Classification techniques with embedded methods

These methods incorporate feature selection directly into the training of the model (Lal et al., 2006), such as with regularisation (Zhang et al., 2020), where certain variable coefficients (the ‘weight’ applied to the feature) are intentionally set to 0 to reduce overfitting. This has the innate effect of eliminating irrelevant features. The following section investigates a set of classifiers with embedded methods applied.

2.8.2. Support vector machines

The SVM model is generated using a randomised grid search with 5-fold cross-validation. This grid search evaluated SVMs with linear, polynomial (degrees 2–5), and radial basis function (RBF) kernels for regularization costs of 1 to 1000 logarithmically. The product of the grid search was an RBF kernel with gamma equal to $1e-4$ and a regularization cost of 1. To perform feature extraction, the permutation importance (Altmann et al., 2010) of the best model in fitting the test data to its labels is evaluated. This allows for the extraction of unbiased feature importance with non-linear kernels.

2.8.3. Decision trees and ensemble learning

Performances with both an individual decision tree (to yield an interpretable diagram) and a random forest (for improved classifier accuracy) are assessed. The random forest is implemented with 50

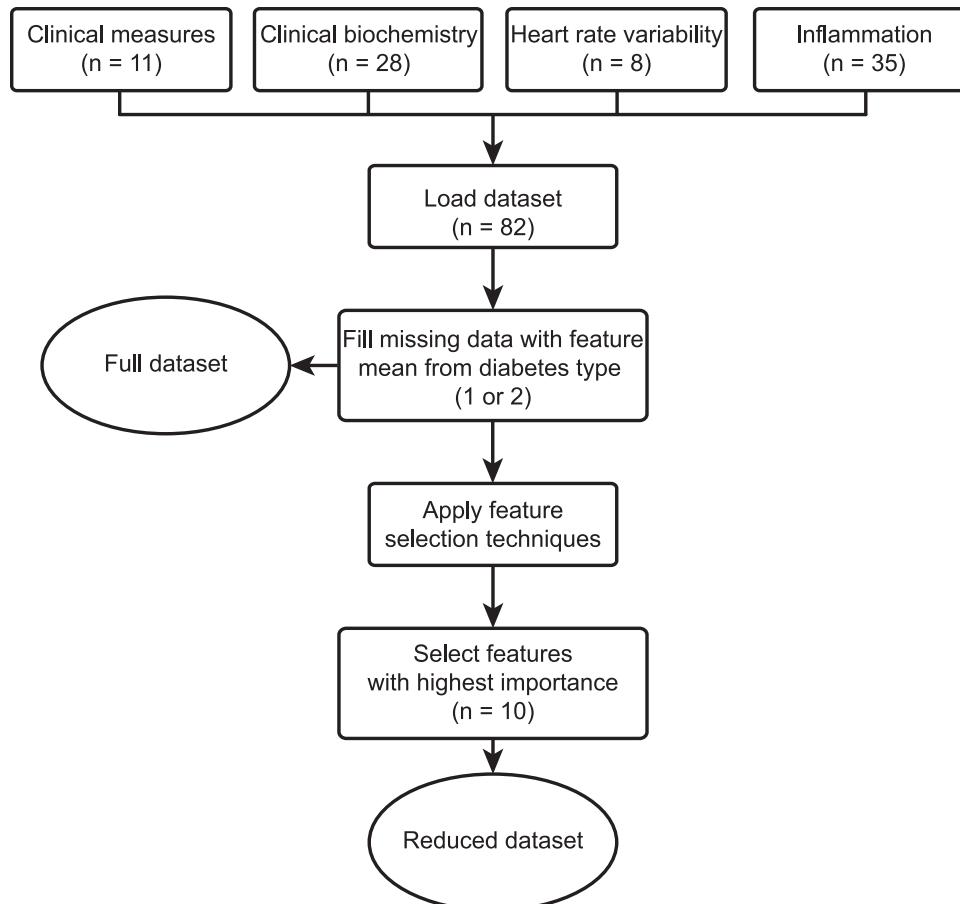


Fig. 1. Pre-Processing pipeline and number of features from each subdomain of the data.

trees with a maximum depth of 8 for minimising overfitting and, like the single decision tree, utilise the mean decrease in impurity to extract the 10 most important features. Importantly, for both the decision tree and random forest classifiers, the comparatively smaller sample size of CAN 1 patients is compensated for by applying class weight balancing, thereby penalising mispredictions of CAN 1 more harshly than CAN 0 or CAN 2. This results in the classifier prioritising the learning of CAN 1 predictions.

2.8.4. Evaluation of classifiers

Two metrics are used for the classifier evaluation: both its test/training accuracy and F1-score. Whilst the former helps with identifying overfitting, the latter is useful for assessing performance with imbalanced classes (Cruz et al., 2016). This is particularly relevant for this dataset, as the proportion of CAN 1 patients is significantly smaller than the other classes.

2.8.5. Feature importance ensembling

For this method, a similar approach to Rajiv Shah et al. (Shah and Peretiatko, 2021) is taken, but instead weighing the scores by model performance, to prioritise features that contribute to a more successful model.

Algorithmically, this is described as follows:

1. Normalise importances for each model (of 10 features) such that they sum to 1
2. Multiply all importances for each model by test performance of model
3. Sum scores for each feature and list in descending order

Which results in the pipeline illustrated in Fig. 2. This creates a set of features that is, effectively, a weighted sum of the best features from each model. This resultingly encourages generalisability of results, as these features should be prevalent irrespective of the classification tool used.

2.8.6. Wrapper methods

Wrapper methods perform feature selection by generating increasingly large subsets (recursively adding features) of the original feature space up to a limit until the best combination of features is found. This has seen medical applications using supervised learning with Naïve Bayes amongst other classification algorithms (Karegowda et al., 2010; Singh and Singh, 2021).

2.8.7. Receiver operator characteristic

Following the extraction of the 10 best features across all models through ensembling, the sensitivity and specificity of the best classifier for a binary comparison of CAN classes were derived using receiver operator characteristic (ROC) curves. A brute force, wrapper approach was implemented by exhaustively generating combinations of variables up to 3 features and evaluating the area under the curve (AUC).

3. Results

3.1. Support vector Machine

The training and test accuracies for the support vector machine are for the full dataset: 92.0% training and 55.0% testing with an F1-score of 53.6%. For the reduced dataset, the accuracies were 59.1% for training and 55.4% for testing with an F1-score of 56.4. The 10 best performing features were: rMSSD, disease duration, SDNN, CVT, VEGF-D, VLF, LF, IL-7, height, and Flt-1. These results are also visualised in Fig. 2. By retraining the model on the 10 most significant features, a similar test accuracy (a 0.4% improvement)

can be achieved with a large reduction ($\approx 33\%$) in training accuracy. This illustrates that overfitting is significantly reduced, allowing for a more generalisable model.

3.2. Decision tree

The training and test accuracies for the decision tree are for the full dataset: 97.0% training and 47.3% testing with an F1-score of 47.1%. For the reduced dataset, the accuracies were 92.7% training and 50.7% testing with an F1-score of 51.8%. The 10 best performing features were: rMSSD, disease duration, HF, VEGF-D SDNN, IL-2, CCL17, age, TSH, and CVT. The four highest decision tree branches are displayed in Fig. 3 and the features are also visualised in Fig. 2. A reduction in training accuracy of $\approx 4.3\%$ permits an increase in test accuracy of $\approx 3.4\%$.

3.3. Random forest

The training and test accuracies for the random forest are for the full dataset: 58.0% training and 61.6% testing with an F1-score of 61.7%. For the reduced dataset, the accuracies were 58.9% training and 59.7% testing with an F1-score of 60.1%. The 10 best performing features were: rMSSD, disease duration, HF, SDNN, CVT, IL-5, VEGF-D, LF, IL-2, and transferrin. These results are also visualised in Fig. 2. The random forest classifier uses the 'out of bag' (OOB) error as a means of quantifying the test and training accuracies. This computes the inaccuracy in predicting a training sample using only the trees that did not have this training sample in their bootstrap. The reduced dataset produced a 0.9% increase in training accuracy with a 1.9% decrease in test accuracy. Due to the similar magnitudes of test and training accuracy, however, it is not the case that this classifier is overfitting. Instead, a relatively high accuracy can be maintained despite the smaller input space.

3.4. Confusion matrices

The accuracies of individual CAN predictions in the confusion matrix of the best performing classifier, which was the random forest classifier, were visualised in Fig. 4. For the full dataset ($n = 90$), the true predictions were: 85.2% (CAN 0), 7% (CAN 1), and 67.5% (CAN 2). For the reduced dataset ($n = 10$), the true predictions were: 76.1% (CAN 0), 18.7% (CAN 1) and 68.2% (CAN 2).

3.5. Feature importance ensembling

Ensembling the features from the three classifiers resulted in the 10 most important features being: rMSSD, disease duration, SDNN, VEGF-D CVT, HF, IL-2, CCL17, LF, and age. Looking at the proportion of total feature importance beat-to-beat data accounted for approximately 60% of the data followed by demographic data and inflammatory markers. Notably, no features from the standard biochemical blood sample dataset were part of the top 10 ensembled features. The 10 most important features were stratified according to the CAN score in Table 2.

3.6. Receiver operator characteristic curves

Evaluating the random forest classifier with binarized classes and performing feature selection with an exhaustive search (up to a maximum of $n = 3$ features) yields the results shown in Supplementary Fig. 1. The extracted features, up to $n = 3$, the sensitivity and specificity at the optimal discriminator (highest AUC of all feature combinations), as well as AUC. These results are summarised in Table 3.

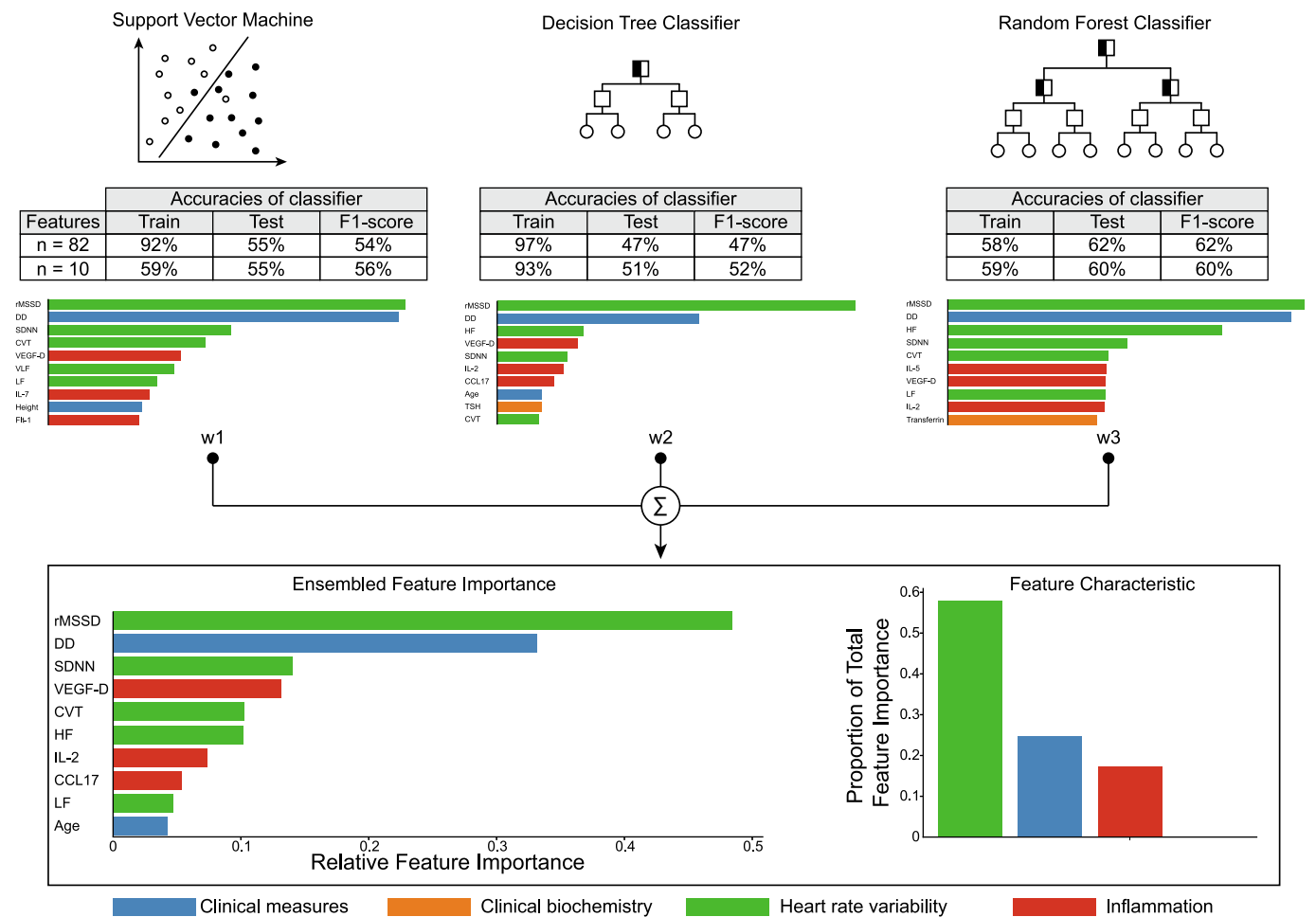


Fig. 2. Training and testing accuracies of the three classifiers support vector machine, decision tree classifier, and random forest classifier in the full (n = 82) and reduced (n = 10) datasets. The method of ensembling takes the best performing features from each classifier accounting for test accuracies. This results in the 10 most important ensembled features. Root mean square of successive differences (rMSSD), disease duration (DD), high frequency (HF), vascular endothelial growth factor (VEGF)-D, standard deviation of the averages of RR (SDNN), Interleukin (IL)-2, -5, CC chemokine ligand (CCL)17, thyroid stimulating hormone (TSH), cardiac vagal tone (CVT), very low frequency (VLF), low frequency (LF), vascular endothelial growth factor receptor 1 (Flt-1).

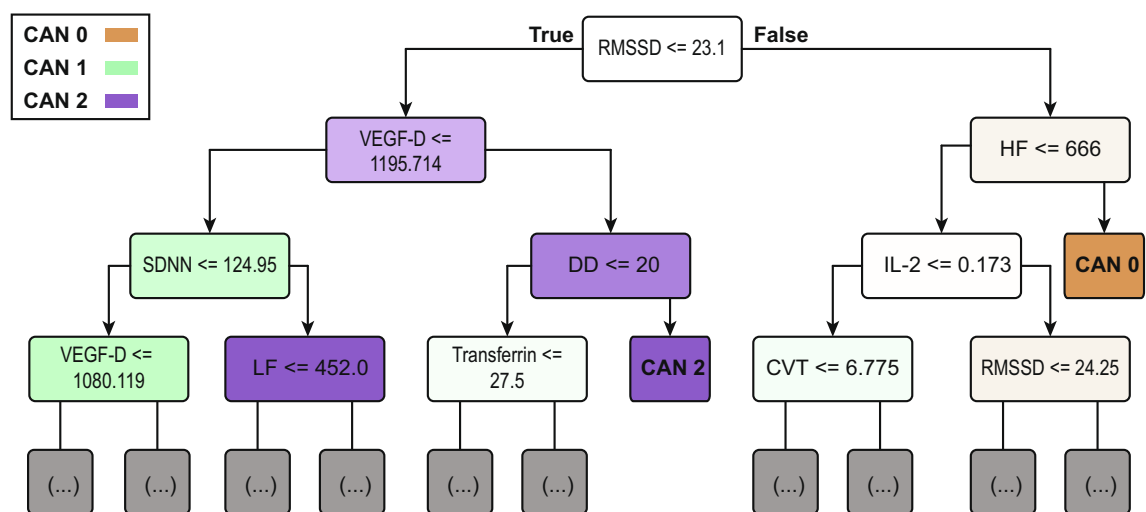


Fig. 3. Structure of an example decision tree. Taking the left path from each node evaluates its rule as true and false otherwise. Darker coloured boxes for each colour correspond to the certainty of prediction. The grey boxes at the bottom of the tree denote truncation, as showing the maximum depth of the tree would hinder readability. Root mean square of successive differences (rMSSD), vascular endothelial growth factor (VEGF)-D, low frequency (LF), standard deviation of the averages of RR (SDNN), disease duration (DD), Interleukin (IL)-2, high frequency (HF), cardiac vagal tone (CVT).

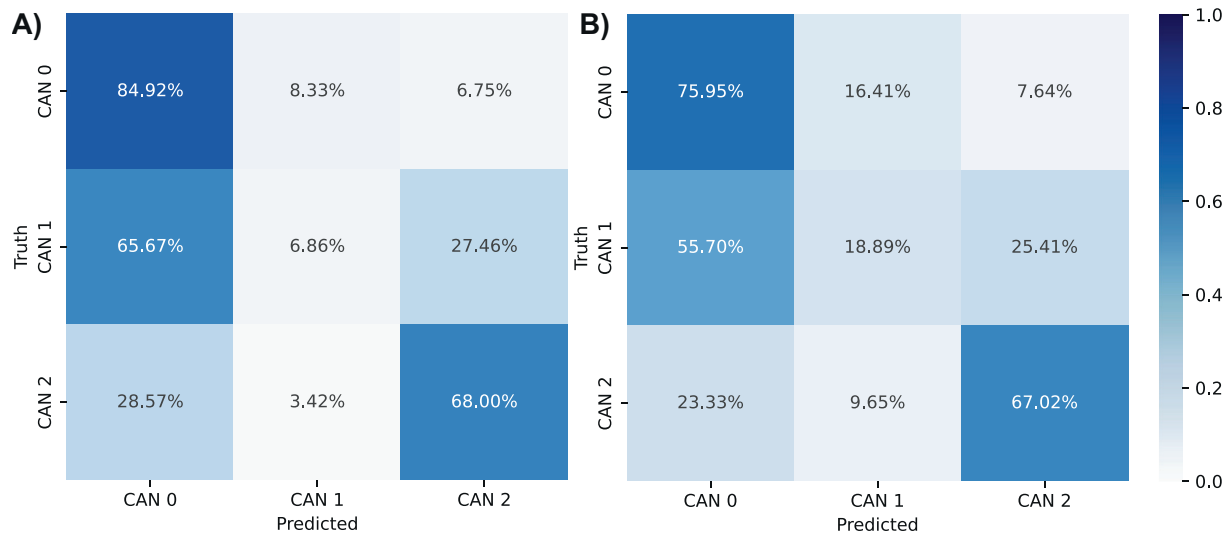


Fig. 4. Class prediction accuracies (confusion) for A) full ($n = 90$) and B) reduced ($n = 10$) dataset based on demographics, blood data, heart data, and inflammation data, and averaged across 1000 random forest classifiers. Each confusion matrix shows the true and predicted accuracy across the left to right diagonal of the cardiovascular autonomic neuropathy (CAN).

Table 2

Ten most important features derived from ensembling and stratified by cardiovascular autonomic neuropathy (CAN) Score. The features are ordered in order of importance. Root mean square of successive differences (rMSSD), standard deviation of the averages of RR (SDNN), vascular endothelial growth factor (VEGF)-D, cardiac vagal tone (CVT), high frequency (HF), Interleukin (IL)-2, CC chemokine ligand (CCL) 17, low frequency (LF).

	CAN Score		
	0	1	2
rMSSD	37.4 ± 17.2	27.2 ± 7.8	21.8 ± 10.1
Disease duration	15.2 ± 11.6	15.8 ± 11.5	27.8 ± 12.3
SDNN	141.3 ± 42.1	115.7 ± 31.1	114.4 ± 32.3
VEGF-D	1266.5 ± 376.4	1380.1 ± 494.8	1756 ± 1029.1
CVT	5.4 ± 3.6	3.7 ± 2	3.6 ± 2.3
HF	474 ± 474.7	217.7 ± 148.6	173.9 ± 166.3
IL-2	0.2 ± 0.1	0.1 ± 0.1	0.2 ± 0.5
CCL17	306.1 ± 200.2	254.6 ± 136.2	469.3 ± 965
LF	949.1 ± 818.2	533.2 ± 433	458.6 ± 363.8
Age	55 ± 17.6	57 ± 13	52 ± 11.3

4. Discussion

Early detection of CAN is clinically relevant, as progression may be halted or even reversed through intervention. To classify the severity of CAN, we identified the 10 most important features using the ensembling method: rMSSD, disease duration, SDNN, VEGF-D, CVT, HF, IL-2, CCL17, LF, and age. Some methods have shown that inflammatory markers and heart rate-derived measures associate with neurocardiac function and pose possible screening tools (Wegeberg et al., 2020a, 2020b; Wegeberg et al., 2022). The variables rMSSD, SDNN, CVT, HF, and LF are all heart rate variability derived time- and frequency domains. Parasympathetic activity affects primarily the time-derived measure RMSSD along with the variability of the high frequency components of frequency data (0.15–0.4 Hz) (Pop-Busui, 2012). These features have all previously been reported to be diminished by CAN, due to parasympathetic withdrawal causing sympathetic dominance (Wegeberg et al., 2020a; Wegeberg et al., 2020b). Disease duration and age were expected to contribute to CAN development throughout diabetic disease progression. Especially in type 2 diabetes, age and disease duration are less interdependent in comparison to type 1 diabetes. VEGF-D, IL-2, and CCL17 are all markers of inflamma-

tion. Increased VEGF-D has been reported to be overexpressed in diabetic atherosclerosis (Wong et al., 2011), which is associated with reduced heart rate variability (Jaiswal et al., 2013). IL-2 has been associated with changes in sympathetic activity, but the subtle magnitude of the marker limits its use as a clinical biomarker (Guinjoan et al., 2009; Tio et al., 2000). CCL17 level is increased with age and correlates with cardiac dysfunction (Zhang et al., 2022).

4.1. Evaluation of classifiers

Although many models exist for performing supervised learning on a clinically obtained dataset, the models in this paper are selected for the following reasons: the SVM is a well-established classifier that has seen countless medical implementations and is known to offer high performance. However, due to its functionality as a ‘black box,’ other methods are preferable, for example if visualising the numerical ‘cut-offs’ that classify a patient is desired. In this case, a decision tree model is preferred. This model is known to suffer from overfitting, so the use of a random forest model is often used to negate this problem, although this comes at the cost of visualisation. Since all models offer positives and negatives to this task, and all may produce varying features, it is important to pool their responses to produce a set of features that persists across all models (ensembling), to maximise generalisability.

4.1.1. Quantitative analysis

The use of a classifier for feature selection is entirely dependent on its performance. With performance ranging from $\approx 51\%$ with decision trees to $\approx 60\%$ with ensembled random forests, we accept that this model is insufficient as a clinical tool if an accurate diagnosis needs to be performed. Instead, this model is offered for explorative purposes. Randomly guessing a patient’s CAN score would yield an accuracy of 33.3%, and “strategically” guessing the CAN score based on the majority class (CAN 0) would yield 45.1% accuracy. For this reason, having demonstrated that the best model offers a 15%-point improvement in classification accuracy, these features may contribute to an understanding of the underlying mechanisms resulting in CAN. The training accuracy is reduced in all cases despite maintaining or increasing the test accuracy, this illustrates that the features selected are suitable for generalising the model, thereby reducing overfitting to the dataset.

Table 3

Summary of information from receiver operator characteristic (ROC) curve analysis. Cardiovascular autonomic neuropathy (CAN), standard deviation of the averages of RR (SDNN), CC chemokine ligand (CCL)17, Interleukin (IL)-2, root mean square of successive differences (rMSSD), vascular endothelial growth factor (VEGF)-D.

CAN	Features	Sensitivity	Specificity	AUC
CAN 0, CAN 1	SDNN, CCL17, Age	0.7	1.0	0.883
CAN 0, CAN 2	IL-2, CCL17, Age	1.0	0.88	0.922
CAN 0, CAN 1 + 2	rMSSD, CVT, Age	0.7	0.95	0.867
CAN 1, CAN 2	Disease Duration, VEGF-D	0.85	0.9	0.946

Generating the confusion matrices for both the full and reduced datasets for the best performing classifier (random forest) allows for a direct comparison of the inter-class classification accuracies. From Fig. 4, both the full and reduced datasets perform well at classifying CAN 0 and CAN 2, with accuracies of 85.2% and 67.5% for the full dataset and 76.1% and 68.2%, for the reduced dataset. This demonstrates that the full dataset is better when categorising CAN 0, offering a $\approx 9\%$ improvement, however, both classifiers offer similar performance with regard to categorising CAN 2.

Both classifiers struggle with the classification of CAN 1 as the full dataset offers a 7% accuracy, with 66.6% of the predictions incorrectly assigned to CAN 0. This may reflect the clinical challenge in the early detection of CAN 1. Similarly, the reduced dataset has an 18.7% classification accuracy for CAN 1, with 55.8% of these predictions assigned to CAN 0. From these results, the difficulty in distinguishing between CAN 1 and CAN 0 is evident, although the reduced dataset does offer a considerable, threefold improvement of $\approx 12\%$. This suggests that feature selection ‘de-noises’ the classifier well.

4.2. Clinical implications

To our surprise, none of the standard clinical biochemical analytes, such as those describing the glycaemic level, lipidic level, or kidney function were included in the most important features accounting for prediction accuracies in any of the three classifiers, nor in the ensemble model. This may be due to a non-linear development of diabetic microvascular complications developing independently of each other. Furthermore, this may underscore the importance of recognizing systemic inflammatory markers in the investigation of CAN and is supported by other findings that associate interleukin markers with dysfunctional heart rate variability (Wegeberg et al., 2022). The clinical classification of CAN 1 is difficult in both the full and reduced dataset, which underlines the complexity of this subsection of the patient population and emphasizes the silent nature and multifaceted underlying pathophysiology. However, our findings implies that cardiac reflex responses should be monitored closely to diagnose and classify severity levels of CAN accurately.

4.3. ROC curve analysis

We used feature selection to find the features that best described the distinction between different CAN stages, but no more than three. Interestingly, different combinations of features appear to provide the best distinction between the different stages, presenting with highly sensitive and specificity scores at the optimal disclination point. As expected, the group with severe CAN (CAN2) were easier to discriminate from no or less severe CAN (CAN 0 and CAN 1), providing higher AUC based on inflammatory markers and age or disease duration, though interestingly, without taking into consideration heart rate variability measures. For distinguishing less severe CAN (CAN 1) from no CAN (CAN 0), which is useful for early detection of cardiovascular autonomic neuropathy, a heart rate variability element was, however, necessary to

achieve optimal discrimination. Additionally, age appears to be a prominent factor when distinguishing no CAN from the presence of any type of CAN, suggesting CAN arises later in life.

4.4. Limitations

The recruitment in the cohorts were vulnerable to selection bias as in particularity individuals with a surplus of resources, might be overrepresented. The choice to not include diabetes type into the model was done to minimise the risk of overfitting data since a large part of the group of people with CAN 2 have type 1 diabetes. We hypothesise that while the underlying differences between diabetes type one and two are very different the effects of inadequate management of the disease will have similar systemic bodily effect. CAN has previously been shown to be independent of diabetes type (Pop-Busui, 2012). In machine learning, the number of features that can be utilised is closely related to the number of subjects that are being analysed. This is observable in the overfitting in the full dataset parts of the analysis, but mitigated by selecting fewer features resulting in closer resemblance in the testing and training accuracies. A partial explanation for the lower classification accuracies of CAN 1 is the class imbalance with CAN 1 having the lowest number of participants (22.55%) and CAN 0 having the largest (45.1%).

5. Conclusion

Surprisingly, none of the standard clinical biochemical analytes, describing the glycaemic level, lipidic level, or kidney function were included in the ten most important features associated with CAN. In contrast, the most important features are beat-to-beat measures accounting for approximately 60% of the feature importance of the ensemble data. Unsurprisingly, the extracted beat-to-beat measures are primarily suggestive of reduced parasympathetic activity. Furthermore, disease duration and age are expected to correlate with CAN development, since these increase with disease progression. Most importantly, our models indicate a higher impact of markers reflecting systemic inflammation. Thus, new treatment modalities targeting systemic low-grade inflammation could potentially serve as a therapeutic target in the attempt to halt the development and progression of CAN.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

No specific funding was used for this publication.

Author statements

The corresponding author, Rasmus Bach Nedergaard, has full access to all data and the right to publish such data. All authors participated in a meaningful way in the preparation of the manuscript.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.clinph.2023.06.011>.

References

- Abdalrada AS, Abawajy J, Al-Quraishi T, Islam SMS. Prediction of cardiac autonomic neuropathy using a machine learning model in patients with diabetes. *Ther Adv Endocrinol Metab* 2022;13:1–10. <https://doi.org/10.1177/20420188221086693>.
- Al Jarullah AA. Decision tree discovery for the diagnosis of type II diabetes. In: 2011 International Conference on Innovations in Information Technology, IEEE; 2011, p. 303–7. <https://doi.org/10.1109/INNOVATIONS.2011.5893838>.
- Altmann A, Tološi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. *Bioinformatics* 2010;26:1340–7. <https://doi.org/10.1093/bioinformatics/btq134>.
- Kumari VA, Chitra R. Classification of diabetes disease using support vector machine. *Int J Eng Res Appl* 2013;3:1797–801.
- Breiman L. Random Forests. *Mach Learn* 2001;45:5–32. <https://doi.org/10.1023/A:1010933404324>.
- Brock C, Hansen CS, Karmisholt J, Møller HJ, Juhl A, Farmer AD, et al. Liraglutide treatment reduced interleukin-6 in adults with type 1 diabetes but did not improve established autonomic or polyneuropathy. *Br J Clin Pharmacol* 2019;85:2512–23. <https://doi.org/10.1111/bcp.14063>.
- Burgess JF, Christiansen CL, Michalak SE, Morris CN. Medical profiling: improving standards and risk adjustments using hierarchical models. *J Health Econ* 2000;19:291–309. [https://doi.org/10.1016/S0167-6296\(99\)00034-X](https://doi.org/10.1016/S0167-6296(99)00034-X).
- Cherkassky V, Dhar S. Interpretation of Black-Box Predictive Models. Measures of Complexity, Cham: Springer International Publishing; 2015, p. 267–86. https://doi.org/10.1007/978-3-319-21852-6_19.
- Chung WK, Erion K, Florez JC, Hattersley AT, Hivert MF, Lee CG, et al. Precision medicine in diabetes: a consensus report from the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care* 2020;43:1617–35. <https://doi.org/10.2337/dci20-0022>.
- Croghan W, Egeghy P. Methods of Dealing with Values Below the Limit of Detection using SAS Carry. St. Petersburg: Southeastern SAS User Group; 2003.
- Cruz R, Fernandes K, Cardoso JS, Pinto Costa JF. Tackling class imbalance with ranking. In: 2016 International Joint Conference on Neural Networks (IJCNN), IEEE; 2016, p. 2182–7. <https://doi.org/10.1109/IJCNN.2016.7727469>.
- Ejskjaer N, Jesper F, Jacob F, Jacobsen PE, Poulsen PL, Nygaard H. A pocket-size device to detect autonomic neuropathy. *J Diabetes Sci Technol* 2008;2:692–6. <https://doi.org/10.1177/193229680800200421>.
- Feldman EL, Callaghan BC, Pop-Busui R, Zochodne DW, Wright DE, Bennett DL, et al. Diabetic neuropathy. *Nat Rev Dis Primers* 2019;5. <https://doi.org/10.1038/s41572-019-0092-1>.
- Fleischer J, Nielsen R, Laugesen E, Nygaard H, Poulsen PL, Ejskjaer N. Self-Monitoring of cardiac autonomic function at home is feasible. *J Diabetes Sci Technol* 2011;5:107–12. <https://doi.org/10.1177/193229681100500115>.
- Guinjoan SM, Vigo DE, Castro MN, Tateosian N, Chuluyan E, Costanzo E, et al. Mood, Th-1/Th-2 cytokine profile, and autonomic activity in older adults with acute/decompensated heart failure: Preliminary observations. *World J Biol Psychiatry* 2009;10:913–8. <https://doi.org/10.1080/15622970802432153>.
- Habibi S, Ahmadi M, Alizadeh S. Type 2 Diabetes Mellitus Screening and Risk Factors Using Decision Tree: Results of Data Mining. *Glob J Health Sci* 2015;7. <https://doi.org/10.5539/gjhs.v7n5p304>.
- Haque F, Reaz MBI, Chowdhury MEH, Malik R, Alhatou M, Kobashi S, et al. A machine learning-based severity prediction tool for diabetic sensorimotor polyneuropathy using Michigan neuropathy screening instrumentations. 2022.
- Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer Science + Business Media, Inc.; 2009.
- Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng* 2007;9:90–5. <https://doi.org/10.1109/MCSE.2007.55>.
- Jaiswal M, Urbina EM, Wadwa RP, Talton JW, D'Agostino RB, Hamman RF, et al. Reduced heart rate variability is associated with increased arterial stiffness in youth with Type 1 Diabetes. *Diabetes Care* 2013;36:2351–8. <https://doi.org/10.2337/dci12-0923>.
- Karegowda AG, Jayaram MA, Manjunath AS. Feature subset selection problem using wrapper approach in supervised learning. *Int J Comput Appl* 2010;1:13–7.
- Lal TN, Chapelle O, Weston J, Elisseeff A. *Embedded Methods. Feature Extraction*, Berlin, Heidelberg: Springer Berlin Heidelberg; 2006, p. 137–65. https://doi.org/10.1007/978-3-540-35488-8_6.
- Ljubic B, Hai AA, Stanojevic M, Diaz W, Polimac D, Pavlovski M, et al. Predicting complications of diabetes mellitus using advanced machine learning algorithms. *J Am Med Inform Assoc* 2020;27:1343–51. <https://doi.org/10.1093/jamia/ocaa120>.
- Malik M, Bigger JT, Camm AJ, Kleiger RE, Malliani A, Moss AJ, et al. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *Eur Heart J* 1996;17:354–81. <https://doi.org/10.1093/oxfordjournals.eurheartj.a014868>.
- Othman MF bin, Abdullah NB, Kamal NFB. MRI brain classification using support vector machine. In: 2011 Fourth International Conference on Modeling, Simulation and Applied Optimization, IEEE; 2011, p. 1–4. <https://doi.org/10.1109/ICMSAO.2011.5775605>.
- Pedregosa F, Varoquaz G, Gramfort A. Scikit-learn: Machine Learning in Python. *Mach. Learn.* 2011;2825–30.
- Pop-Busui R. What do we know and we do not know about cardiovascular autonomic neuropathy in diabetes. *J Cardiovasc Transl Res* 2012;5:463–78. <https://doi.org/10.1007/s12265-012-9367-6>.
- Shah R, Peretiak V. Using Feature Importance Rank Ensembling (FIRE) for Advanced Feature Selection; 2021. <https://www.datarobot.com/blog/using-feature-importance-rank-ensembling-fire-for-advanced-feature-selection/>.
- Seijo-Pardo B, Bolón-Canedo V, Porto-Díaz I, Alonso-Betanzos A. Ensemble Feature Selection for Rankings of Features; 2015, p. 29–42. https://doi.org/10.1007/978-3-319-19222-2_3.
- Singh N, Singh P. A hybrid ensemble-filter wrapper feature selection approach for medical data classification. *Chemometr Intell Lab Syst* 2021;217. <https://doi.org/10.1016/j.chemolab.2021.104396>.
- Sinreich M, Taylor BV, Dyck PJ. Diabetic Neuropathies. *Neurologist* 2005;11:63–79. <https://doi.org/10.1097/01.nrl.0000156314.24508.ed>.
- Spallone V, Bellavere F, Scionti L, Maule S, Quadri R, Bax G, et al. Recommendations for the use of cardiovascular tests in diagnosing diabetic autonomic neuropathy. *Nutr Metab Cardiovasc Dis* 2011;21:69–78. <https://doi.org/10.1016/j.numecd.2010.07.005>.
- Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Res Clin Pract* 2022;183. <https://doi.org/10.1016/j.diabres.2021.109119>.
- Tio RA, Niekne J, de Vries EGE, Pfeiffer C, de Jongste MJL, Pieper E, et al. Negative inotropic effects of recombinant interleukin 2 in patients without left ventricular dysfunction. *Eur J Heart Fail* 2000;2:167–73. [https://doi.org/10.1016/S1388-9842\(00\)00073-8](https://doi.org/10.1016/S1388-9842(00)00073-8).
- Vinik AI, Erbas T, Casellini CM. Diabetic cardiac autonomic neuropathy, inflammation and cardiovascular disease. *J Diabetes Investig* 2013;4:4–18. <https://doi.org/10.1111/jdi.12042>.
- Wang F, Gatsonis CA. Hierarchical models for ROC curve summary measures: design and analysis of multi-reader, multi-modality studies of medical tests. *Stat Med* 2008;27:243–56. <https://doi.org/10.1002/sim.2828>.
- Wegeberg AM, Lunde ED, Riahi S, Ejskjaer N, Drewes AM, Brock B, et al. Cardiac vagal tone as a novel screening tool to recognize asymptomatic cardiovascular autonomic neuropathy: aspects of utility in type 1 diabetes. *Diabetes Res Clin Pract* 2020a;170. <https://doi.org/10.1016/j.diabres.2020.108517>.
- Wegeberg AM, Okdahl T, Riahi S, Ejskjaer N, Pociot F, Størling J, et al. Elevated levels of interleukin-12/23p40 may serve as a potential indicator of dysfunctional heart rate variability in type 2 diabetes. *Cardiovasc Diabetol* 2022;21:1–10. <https://doi.org/10.1186/s12933-021-01437-w>.
- Wegeberg AM, Okdahl T, Fløyel T, Brock C, Ejskjaer N, Riahi S, et al. circulating inflammatory markers are inversely associated with heart rate variability measures in Type 1 Diabetes. *Mediators Inflamm* 2020. <https://doi.org/10.1155/2020/3590389>.
- Wong BW, Wong D, Luo H, McManus BM. Vascular endothelial growth factor-D is overexpressed in human cardiac allograft vasculopathy and diabetic atherosclerosis and induces endothelial permeability to low-density lipoproteins in vitro. *J Heart Lung Transplant* 2011;30:955–62. <https://doi.org/10.1016/j.healun.2011.04.007>.
- Xu W, Zhang J, Zhang Q, Wei X. Risk prediction of type II diabetes based on random forest model. In: 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), IEEE; 2017, p. 382–6. <https://doi.org/10.1109/AEEICB.2017.7972337>.
- Zhang H, Wang J, Sun Z, Zurada JM, Pal NR. Feature selection for neural networks using group lasso regularization. *IEEE Trans Knowl Data Eng* 2020;32:659–73. <https://doi.org/10.1109/TKDE.2019.2893266>.
- Zhang Y, Ye Y, Tang X, Wang H, Tanaka T, Tian R, et al. CCL17 acts as a novel therapeutic target in pathological cardiac hypertrophy and heart failure. *J Exp Med*. 2022;219:e20200418.