Informatics in Medicine Unlocked 17 (2019) 100266

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: http://www.elsevier.com/locate/imu

Analysis of Friedreich's ataxia patient clinical data reveals importance of accurate GAA repeat determination in disease prognosis and gender differences in cardiac measures

Mohammadmersad Ghorbani [a], Françoise Pousset [b], Allan Tucker [a], Stephen Swift [a], Paola Giunti [c], Michael Parkinson [c], David Gilbert [a], XiaoHui Liu [a], Annette Payne [a,*], the Clinical Partners of the EFACTS Consortium

[a] Dept. of Computer Science, College of Engineering, Design and Physical Brunel University London, UK
[b] Département de Cardiologie, La Pitié Salpètrière, Paris, France
[c] Department of Molecular Neuroscience, Institute of Neurology, UCL, London, UK

ARTICLE INFO

Keywords:
Friedrich's ataxia
Nucleotide repeat
Computerised classification
Neurology
Clinical prognosis
Intelligent data analysis

ABSTRACT

Friedreich's ataxia (FRDA) is a rare autosomal recessive inherited neurodegenerative disease which is the result of a triplet repeat expansion in the intronic region of the frataxin *FXN* gene resulting in depleted frataxin protein expression. Disease onset is usually in childhood and causes progressive damage to the nervous system resulting in progressive disability. This work uses computer aided classification techniques to identify which measures of the disease progression, including accurate determination of the shortest allele repeat length, are the most informative when trying to predict likely disease progression and prognosis. Further we investigate the possibility of a gender difference in the progression of the disease. Our results highlight the importance of accurate determination GAA repeat length in any clinical predictions showing that the number of repeats is the best prognostic tool in FRDA and is strongly linked to the age at onset disease. Further that there are possible gender dependent differences in cardiac measurements recorded from patients of similar age of onset and GAA repeat length.

1. Background

Friedreich's ataxia (FRDA) is a rare autosomal recessive inherited neurodegenerative disease with an onset usually in childhood. The disease causes progressive damage to the nervous system resulting in progressive disability. Patients present with initial symptoms of poor coordination such as gait disturbance. Other symptoms can include any of the following: muscle weakness in the arms and legs, visual impairment, hearing impairment, slurred speech, curvature of the spine (scoliosis), high plantar arches (pes cavus deformity of the foot), diabetes and heart disorders (e.g., hypertrophic cardiomyopathy and associated arrhythmias). Onset of symptoms is usually between the ages of 5 and 15 years, but may occur above the age of 25 which is referred to as the late onset form [1–4]. The age of onset has a profound effect on the disease severity and progression with a faster deterioration associated with a younger age of onset [5]. Most of the non neurological features are more common in the typical-onset patients compared to the late-onset [6].

The disease is typically caused by an expansion of an intronic GAA triplet repeat in the *FXN* gene leading to reduced expression of the mitochondrial protein frataxin [7]. The length of the expansion on the shorter allele of the FXN locus is associated with age of onset of disease, with a younger age being associated with a longer repeat [5]. Logistic regression models for the presence of symptoms have demonstrated that the value of GAA repeat length alone of the shorter allele is the best prognosis indicator of the disease and the age at onset, as well as severity of ataxia signs, sex, and presence of neonatal problems [6]. The disease is also associated with epigenetic changes including aberrant DNA methylation leading to downregulation of gene expression [8].

Disease progression is typically monitored by a variety of clinical bedside tests [9–12]. Disease progression, as measured by the Scale for the Assessment and Rating of Ataxia (SARA) base line scale, is more

* Corresponding author.
E-mail address: annette.payne@brunel.ac.uk (A. Payne).

https://doi.org/10.1016/j.imu.2019.100266
Received 19 July 2019; Received in revised form 11 November 2019; Accepted 11 November 2019
Available online 19 November 2019

rapid in patients with a longer GAA repeat on the shorter allele of the FXN locus. Other measurements that are commonly taken are the Inventory of Non-Ataxia Signs (INAS), the Spinocerebellar Ataxia Functional Index (SCAFI), the quality of life measures activities of daily living (ADL) and EQ-5D-3L index [5]. All of these various tests along with the SARA test attempt to assess the degree of ataxia through measurements of the ability to walk and carry out other motor functions, show similar deterioration as the SARA scale as the disease progresses. To date however all the measures have been considered singlely and not in combination with each other or with other measures that may influence disease severity or progression. This study attempts to explore several possibilities within the data that can only be uncovered by complex computational techniques. Firstly we wished to explore if a given combination of measures taken together could more accurately predict severity or progression. Further, given that the SARA score has been shown to be a very accurate measure for disease progression, whether it could be used to predict the size of the repeat expansion of the shorter allele given that measurement of this expansion can be technically problematic and the expansion may change over a person's life time. Also we wished to assign more accurate cutoff ranges for the classification of patients into groups which may assist the selection of patients for clinical trials and to identify the key features that could signify progression points of the disease. Lastly explore the suggestion that gender may exert an influence which to date has not been explored. Thus the overarching objective of this study was to assess which tests best capture disease severity and progression within a large prospective natural history study.

Clinical and demographic data were collected by the clinical partners of the European Friedreich's Ataxia Consortium for Translational Studies (EFACTS) from over 600 genetically confirmed FRDA patients. Ethical approval for the collection of the data was given as part of the separate clinical study and this study has been published previously [5]. In summary the measurements for 61 clinical features were collected with informed consent from the patients. The primary outcome measure was the Scale for the Assessment and Rating of Ataxia measurements (SARA). The secondary outcome measures were: the Inventory of Non-Ataxic Symptoms (INAS); the performance-based coordination test, the Spinocerebellar Ataxia Functional Index (SCAFI); the neurocognitive phonemic verbal fluency test; and two quality-of-life measures: the EQ-5D; and the activities of daily living (ADL), part of the Friedreich's Ataxia Rating Scale (FARS). Further demographic, cardiac, family history, genetic and personal data were collected from each patient.

To date complex computational methods have not been employed to analyse ataxia patient data before with the challenge that this data has been collected from disparate sites across Europe by different clinical teams raising the possibility of inter-clinic variability of measurements. This work was therefore undertaken to not only discover new findings about the disease but also to valid the techniques for use on data with such potential uncontrolled variability. The experiments were run on data from single clinics first before using data from across all the clinics to test the consistency of the results when inter-clinic variation was possible in the data. The computational data analysis techniques where specifically used to identify which features are the most informative in classifying patients into subgroups commonly used to assist in prognosis by clinicians, identifying key stages in disease progression and predicting disease progression in patient subgroups.

2. Methods

The Hill Climbing algorithm was used to optimise the selection of the most informative features based on their ability to discriminate between the states you wish to explore in a given experiment. The algorithm is iterative and is a local search optimisation technique. It will start with an arbitrary solution to the problem (in this case finding the most informative features) then by making an incremental change it attempts to find a better solution. If the algorithm is successful another change is made and so on until no further improvement to the solution can be found. Thus for each experiment a different set of optimum features were discovered based on the aim of experiment which were then used in the machine learning phase. The Hill Climbing algorithm used is given below.

The computational analysis was then carried out using supervised machine learning [13]. We selected classification as the method of choice as it allowed us to predetermine how we would like the patients to be subdivided. These subdivisions being the most clinically useful based on current clinical practice. These classification techniques take raw data and determine which features (in this case features and measures for each patient held in the EFACTS database) enable individuals to be classified into predetermined categories or groups. These methods also allow an assessment of how accurate the data analysis is at generating the categories based on the items the computer selected. Accuracy of the generated model was determined by the percentage individuals correctly classified through a Kappa score [14]. Higher Kappa scores during a classification experiment indicate increased likelihood of identifying attributes which are predicted to categorise patients correctly. Kappa is a measure of the predictive performance of the classifiers, as it takes into account correct classifications that can happen by chance. The highest possible Kappa score that could be generated is 1.0.

Several algorithms exist for the classification of data [13] 37 different machine learning algorithms were tried including Bayes Networks, Naive Bayes, Random Forest, LAD, Tree LMT as part of these experiments. In this study the three most common (Support Vector Machine (SVM), logistical and J48) were the most informative when assessed. Within supervised machine learning paradigms, it is usual to trial more than one algorithm on each data set, as different algorithms have been shown to work better on different data sets [15]. The best method was determined empirically by experimentation. In this study, the Decision Tree inferred using C4.5 [16] was found to give more informative insight when evaluated using 10-fold cross-validation. The algorithms implemented using the WEKA software suite [17], which is an open source package written in Java. Before uploading the data into the software, any individuals with missing values were excluded. All results were presented as the percentages that correctly classified along with the Kappa score. To aid understanding for those unfamiliar with the analytic technique, the decision tree was visualised. The decision tree graphically illustrates the attributes the computer identified as important in correctly classifying the individuals and how they interrelate. WEKA has been used to assist the diagnosis and prognosis of many diseases and disorders since its development in 2009 [18]. Examples include Leukaemia [19], Breast Cancer [20], and Type 2 Diabetes [21]. Thus the C4.5 algorithm was used to identify which attributes are the most important to classify the patients into subgroups such as age at onset groups, male and female, and GAA repeat length groups.

3. Results

In the first experiment we sought to classify the patients in to 3 subgroups according to their age of onset a) 0–25 years old, b) 26–39 years old and c) 40 years and over. These subgroups are what have been traditionally used to classify patients by clinicians in practice and used to determined probable prognosis. We determined that the strongest classifier of age at onset disease subgroups is the GAA repeat number (the number of GAA repeats present in a patient's genome at the FRDA locus). This is shown in Fig. 1 where nearly 89% were correctly classified with a kappa of 0.61 (Fig. 1). Furthermore, in any analysis removing this feature decreased the value of kappa (a measure of the accuracy of classification) very significantly. Since the results showed that the GAA repeat length was such a strong classifier we sought to classify patients into groups based on GAA repeat length alone with the algorithm determining which features are the best to predict GAA repeat length. We found an age at onset is almost the only feature needed to determine

Hill Climbing

Input : Number of iteration

1. Initialise Random solution (Solution) for the problem and Random Fitness for the starting point (Fitness) these are sequence of zero and ones which code for random feature subset. 1 shows feature included 0 shows feature not included

1. **FOR** number of iterations

2. Choose random point near Solution (NewSolution) And calculate NewFitness

3. **IF** NewFitness (kappa value for classification method on new feature subset) is better than Fitness

4. Solution = newSolution

5. Fitness = NewFitness

6. **EndIF**

7. **EndFor**

Output : Solution (sequence of zero and ones decoded as a feature subset)

the GAA repeat length. With an age of onset below 18 predicting the repeat length with high accuracy if the GAA repeat is more than 500bp (Fig. 2). This indicates the importance of accurate measurement of the number of repeats as a prognosis tool above any other measure including the SARA measurements, and confirms that the number of repeats is the best prognostic tool in FRDA and is strongly linked to the age at onset disease.

For the next experiment we sought to identify the key features that could signify progression points of the disease. The patients were manually subdivided into 4 groups according to age of onset: a. 0–5, b. 6–25, c. 26–39, d. Over 40 years. The features that were important as markers of disease progression were then identified by classifying which features were most informative at particular time points after the onset of the disease (disease duration). These time points were set at i. 0–20, ii. 20–30, ii. >30 years after onset. This was done by executing multiple classification experiments using data from patients in each age of onset group to attempt to classify them into one of disease duration groups, thereby identifying which parameters are better classifiers. There was not enough data points to carry out the analyses using the over 40 age of onset group. The other groups' data identified features with average of 76% accuracy (Kappa 0.4217) that could subdivide these groups into the three disease duration groupings. These were elements of the SARA score, the cardiac ejection fraction and activities of daily living, with less importance on INAS. Interestingly, cardiac features are not informative as indicators in patients with an age at onset above 26 years indicating that these features do not change significantly with time in these patients. The converse was true for patients with an onset of 5 years or younger.

To establish if there is a difference in severity of male and female patients with similar ages of onset and repeat expansions, and when taking into account all the features, we attempted to classify patients according to gender. This experiment was suggested after a preliminary examination of a subset of the data: All the features were examined for their difference in male and female patients using regression analysis and classification methodologies. The subset of data consisted of 80 patients based in the UK and all seen at the same London clinic. We found no significant difference in any of the features between male and females using statistical analysis except ejection fraction with a p value of 0.002 and possibly left ventricular posterior wall thickness at diastole (LVPWd) with a p value of 0.054. On regression analysis the data also showed that female patients retain a higher ejection fraction as the disease progresses than males. These results are shown in Table 1 and Fig. 3a and b and c.

This led us to hypothesis that male and female patients could be classified successfully if a subset of cardiac features were used. Using the full cohort of patients in the EFACTS study we found that males and females can be correctly classified by gender 69.4% of the time, indicating a possible difference in how the disease affects males and females however the kappa score is only 0.3. It was indeed the cardiac features which classified the patients most accurately (Fig. 4). As a consequence of our analysis we can show that cardiac features are less severe in females than males with a classification accuracy of 65.3%, with left

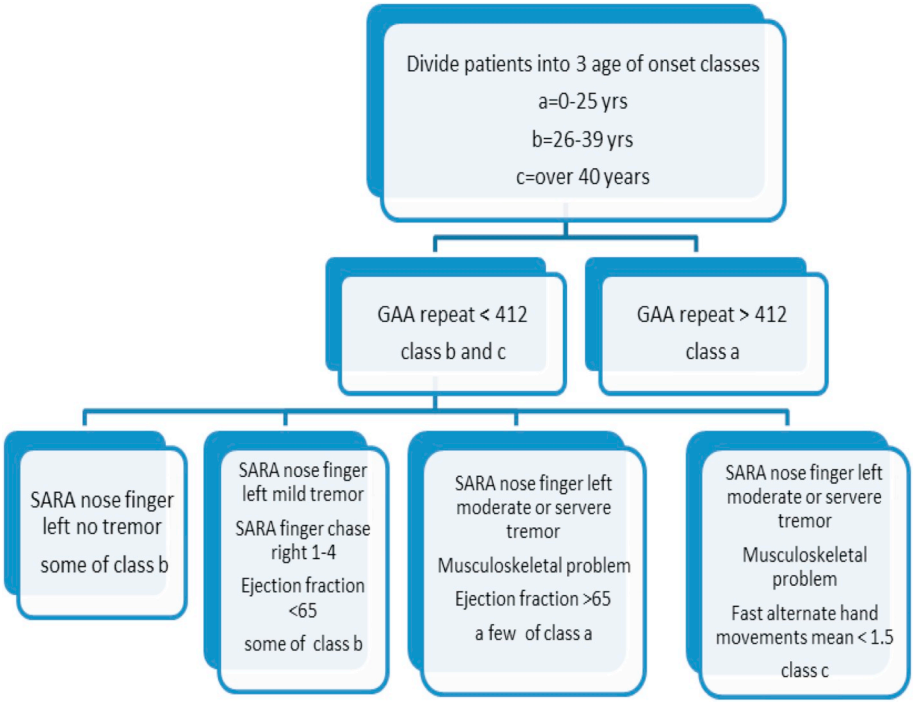

Fig. 1. Decision tree showing which features correctly classifying patients into the following "age at onset" groups: early, middle and late (88.7% correctly classified, kappa = 0.61). When considering ejection fraction as a discriminator a result of between 60 and 65% is considered normal clinically however the algorithm considers 65% to be a good classifier of the groups with fewer GAA repeats. However it should be noted that altered left ventricular ejection fraction is of worse prognosis and can occur in patients with smaller GAA; the impact of which may more important for group b.

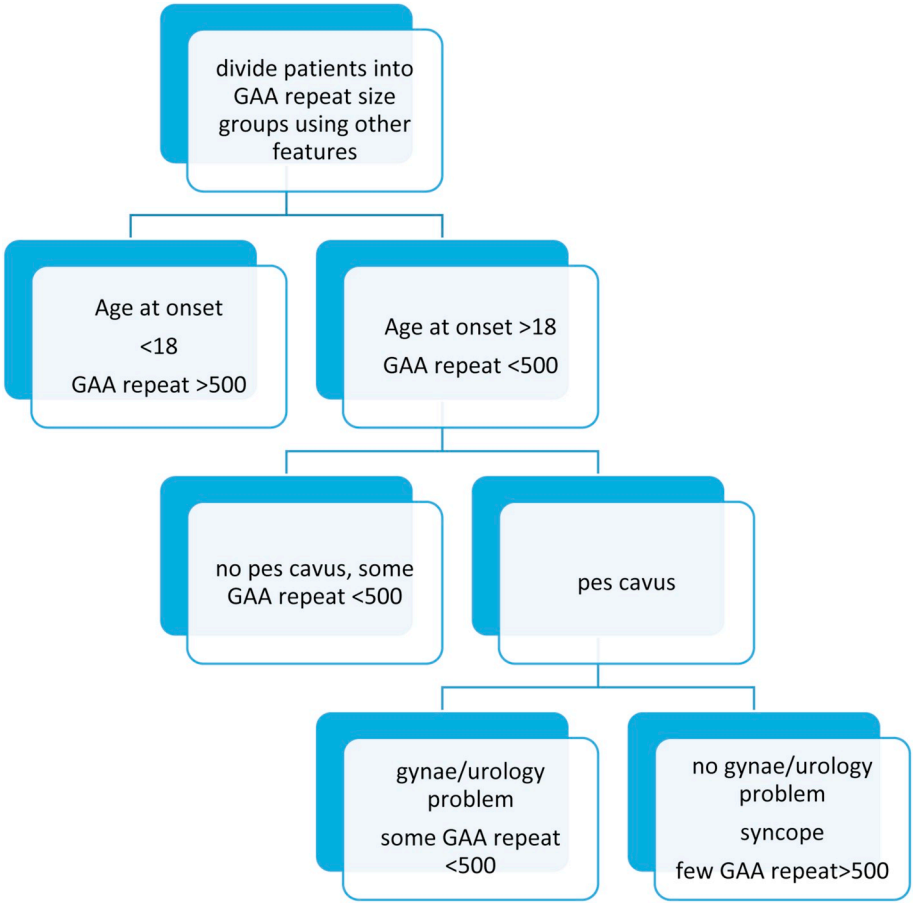

Fig. 2. Decision tree showing which features successfully classify patients into groups of similar GAA repeat size (82.2% correctly classified, kappa = 0.56). This affirms the results of the previous experiment in that age of onset can be used as the strongest indicator of GAA repeat length. Other features indicated on the decision tree whilst useful are far less informative and are only secondary classifiers. Syncope is recorded as a binary integer with relatively few patients reporting this symptom, likewise gynae or urological problems, and neither is further described in the data set so for those patients who have suffered these, the cause is unknown and it may not be related to the FRDA. Pes cavus however is noted to be a symptom of FRDA and therefore is more notable as a classifier in this experiment.

ventricular posterior wall thickness at diastole (LVPWd) the best classifier, usually being less than 10.3 mm in females. The classification results also showed that female patients are more likely to retain sinus rhythm and a larger ejection fraction.

Table 1
Statistical analysis of ejection fraction and LVPWd by gender.

	Male	Female	p-value
Ejection fraction mean	58.48065	66.72632	0.002
Ejection fraction SD	12.40368	9.089129	
LVPWd mean	10.5625	9.47381	0.054
LVPWd SD	2.450306	2.536471	

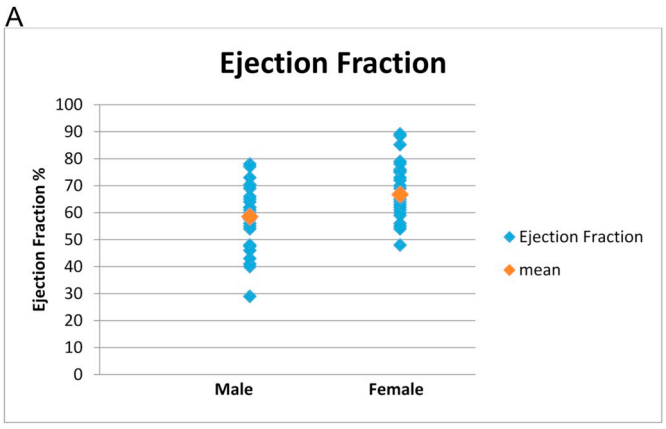

Fig. 3a. Ejection fraction by gender showing the mean for each.

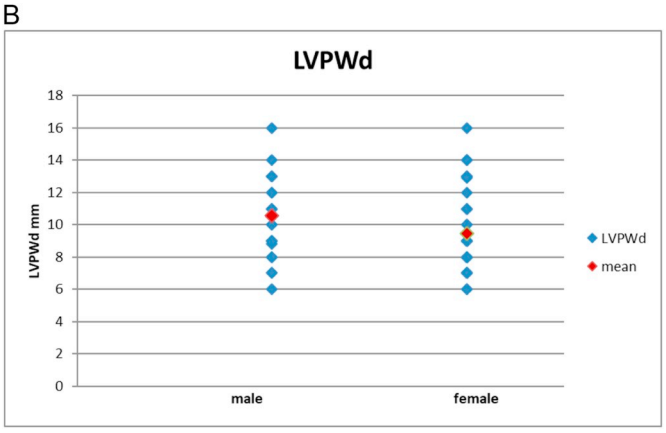

Fig. 3b. LVPWd by gender showing the mean for each.

4. Discussion

It is known that the expansion of the intronic GAA triplet repeat on the shorter allele of the FXN locus is associated with age of onset of disease, with a younger age being associated with a longer repeat and greater disease severity. Further the disease progression rate of patients has long been linked to age at onset which, until accurate GAA repeat length could be determined, was left to clinical observations. The ability to show a more exacting correlation between these has been shown using regression and statistical analyses [5]. Thus our analysis agrees with these previous results. What has not been shown before is just how much more informative the GAA repeat length measure is than any other measurement taken by clinicians to determine the prognosis of a patient. Our results stress the importance of very accurately determining GAA repeat length in patients on diagnosis to more precisely determine possible age of onset, disease severity and prognosis, the other measures being subsidiary to this. GAA repeat length may in the future also be important in identifying which patients are more likely to respond to a treatment and determine when treatments should begin so as to halt the neurodegeneration. The results also suggest that some of the huge

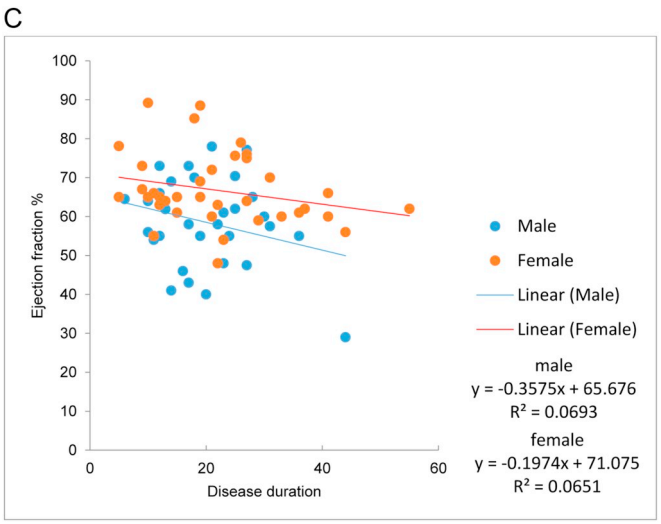

Fig. 3c. Correlation between Ejection Fraction and disease duration showing that ejection fraction decreases with length of time the patient has suffered with the disease. This has been observed previously in previous studies [1]. The novel finding is that this decrease is less pronounced in female patients compared to males. The regression correlation for each gender is shown along with the intersect with the y axis. (The slope for males is −0.3575 compared to −0.1974 for famles and the intersect is 65.676 for males and 71.075 for females (also shown in the figure)). Pearson Correlation Coefficient was calculated to be very weak for the genders: Female R = −0.2694, P = 0.09 and for males R = −0.2632, P = 0.15.

battery of tests that patients undergo, are less informative when trying to predict progression, with only some elements of the SARA test, activities of daily living and some cardiac measures affording any discriminatory value for classifying patients from a computational analysis point of view. Thus we suggest that some tests which we have been shown to be of lesser value could be omitted from the monitoring process. This agrees strongly with the work of the EFACTs consortium who considered each measure individually [5]. We considered in our analyses that some measures, particularly those that are patient self-reported could be less informative in part due to the subjective nature of some of these measures and inter-rater variation, however the data in this study was carefully cross controlled for this. We acknowledge however that such subjective data is not as reliable as the objective measures due to its subjectiveness rather than its ability to inform.

In attempting to discover what features best distinguished progression points of the disease we identified that SARA score, the cardiac ejection fraction and activities of daily living when combined were the most informative features to monitor progression, with less importance on INAS. This agrees strongly with the work of Reetz et al. which looked at each of these measures singly [5]. Interestingly we additionally found that cardiac features are not informative as indicators in patients with an age at onset above 26 years indicating that these features do not change significantly with time in these patients mirroring the less severe disease phenotype seen in these patients. The converse was true for patients with an onset of 5 years or younger who suffer a more severe prognosis. Since the GAA repeat length can change over time in a patient we suggest that a study monitoring changes in the features which best monitor progression should be conducted alongside ongoing monitoring of the repeat length, to establish if any changes in repeat length have a significant effect on progression or not, or if it is only the length at a very early age that is important.

Since there is a suggestion that male and female patients react differently to carrying a similar length deleterious GAA repeat expansion (personal communication from the clinicians of the EFACTS consortium), with women showing less severe symptoms, we sought to determine if male and female patients could be successfully classified

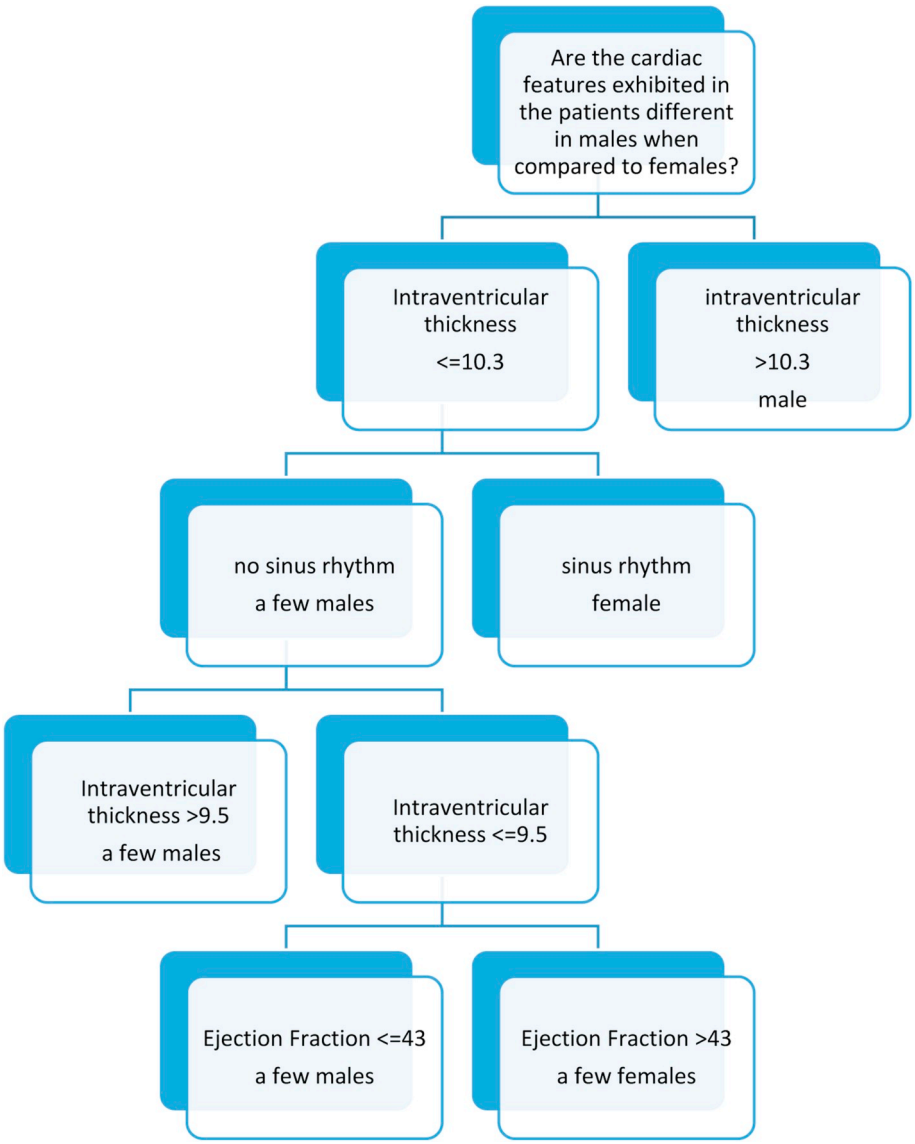

Fig. 4. Decision tree showing which features classify patients according to gender (65.3% correctly classified, kappa = 0.30).

with any accuracy and using which classifiers. It was interesting that we were able to show that they could be correctly classified using cardiac measures, most notably left ventricular thickness, sinus rhythm and ejection fraction all be it with a kappa score of 0.30.

Of the cardiac features included by the computational analysis in the decision tree, three are known to have a prognostic value namely a) sinus rhythm, b) hypertrophy and c) ejection fraction.

a) Sinus rhythm (no sinus rhythm meaning atrial fibrillation) which is used to indicate disease severity.
b) Hypertrophy (cardiac wall thickness) which is highly dependent on the size of the shorter GAA repeat. It has been shown that most FRDA patients are known to have a moderate cardiac concentric hypertrophy (increased intraventricular septal and posterior wall thickness). Moreover in a study by Pousset [22] there was a good correlation between septal wall thickness assessed by echo and MRI where hypertrophy was defined according to the value of septal wall thickness or of left ventricular mass. Further it was been shown that posterior wall thickness tended to decrease over time [22], the evolution of septal wall thickness however remains unclear.
c) Left ventricular ejection fraction (contractility) which in FRDA, cardiac involvement may progress to LV systolic dysfunction and heart failure. In FRDA cardiac function deteriorates slightly with time, but only some patients develop significant alterations of LVEF. The discrimination of the groups at 43% is a significant alteration of LV ejection fraction. When a patient had "no sinus rhythm" the LVEF could decrease however why the hypertrophy (intraventricular septal thickness) is a classifier in the decision tree we cannot explain at this time.

Furthermore, the data suggests female patients show less severe cardiac features as their disease progresses. The majority (53%) of all patients studied in the EFACTs study showed a decrease in ejection fraction over time, however the women less so than the men. It is therefore possible that female patients maintain their near normal cardiac function for longer as the disease progresses due to the possible protective effect of oestrogen [23]. FRDA patients are known to have oxidative stress [24] and there is evidence that oestrogen exerts a protective effect in FRDA patients by reducing oxidative stress in mitochondria. In so doing the mitochondria in females are more able to generate the required amounts of ATP to sustain muscle function [25]. Thus this may go some way to explain a possible gender difference in FRDA patients.

Studies have shown that there are gender differences in cardiac

features e.g. Sandstede et al. [26] and so our data may just reflect maintenance of this difference. Additionally with increased age, males in the healthy population show a significant decrease in volume and mass indices for both heart ventricles, while female values remained unchanged [27]. Ventricular thickness is a well know measure of the risk of heart failure as identified in a study by Spirito et al. [28]. However in FRDA the thickening of the wall of the ventricle does not compensate the alteration of the contractility. We rather observed a thinning of the wall, when the size of the heart increases and the ejection fraction decreases (Pousset, personal communication).

Cardiomyopathy, usually in its moderate hypertrophic form may progress to LV systolic dysfunction and heart failure in FRDA patients. In FRDA cardiac function deteriorates slightly with time, but only some patients develop significant alterations of LVEF. Its presence decreases life expectancy explicitly despite moderate clinical symptoms. The neurological deficit (such as ataxia or muscle weakness) shows no clear relationship with the degree of cardiac involvement [29 and references therein]. Further it has be shown that the extent of left ventricular hypertrophy increases correspondingly with the size of the GAA expansion in 20% of patients in an earlier study [30]. A more recent study by Peverill et al. [31] investigated the effects of FRDA on regional long axis function of the left and right ventricles, and also the relationship of long axis systolic (s') and early diastolic (e') peak velocities with GAA repeat number of the FXN alleles finding all the regional LV s' and e', and both RV s' and RV e', were lower in individuals with FRDA compared to controls. Also the LV septal wall thickness (SWT), RV s' and RV e' were both inversely correlated with the shorter allele repeat, but not with the longer one. Whereas for anterior and lateral s' were the reverse relationships with the allele lengths. Additionally age correlated inversely with e' but not s'. Longer repeat expansion has also been correlated with lower cardiomyocyte counts in patients further linking the disease with cardiac abnormalities [32]. In frataxin conditional knockout mice, cardiac hypertrophy is always followed by dilated cardiomyopathy and heart failure [33]. Recently the presence of cardiovascular disease has been shown to be predicted by sex (male odds ratio = 1.688, CI: 1.280–2.423) by Reetz et al. [6]. Thus the finding that females in our study suffer less hypertrophy than the males is notable.

Some echocardiographic studies show a higher LVEF (left ventricular ejection fraction) in women than in men, but some magnetic resonance imaging (MRI) studies have found comparable LVEFs between genders. [34 and references therein]. It is interesting that these differences are maintained during disease progression in our data. The larger ejection fraction in women therefore could offer some protective effect to those with the disease and prevent thickening of the ventricles seen more prominently in men, which are developed to compensate for cardiac insufficiency. This is a particularly important finding since cardiac complications are a significant cause of mortality in these patients [35].

In summary we believe that computational techniques that perform complex analyses can provide insights into large medical data sets such has been collected in this study. Here we have shown that some data features (shown in the results) are more informative than others in aiding the prognosis and charting the progression of FRDA. We have also shown that there is possibly a difference in the way male and female suffers react to the disease. In the future we hope to employ further techniques to provide additional insights into this data set.

Ethical statement

The authors disclose that there are no financial and personal relationships with other people or organizations that could inappropriately influence (bias) their work.

Declaration of competing interest

We declare that there is no conflict of interest for the manuscript entitled.

Acknowledgements

We would like to thank Brunel University and its staff for its support in kind and all the patients and controls for their participation in the study. Many of the authors were funded as part of EFACTS, which was funded by an FP7 grant from the European Commission (HEALTH-F22010-242193).

The authors declare no conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.imu.2019.100266.

References

[1] Parkinson MH, Boesch S, Nachbauer W, Mariotti C, Giunti P. Clinical features of Friedreich's ataxia: classical and atypical phenotypes. J Neurochem 2013 Aug;126 (Suppl 1):103–17.
[2] Dürr A, Cossee M, Agid Y, Campuzano V, Mignard C, Penet C, et al. Clinical and genetic abnormalities in patients with Friedreich's ataxia. N Engl J Med 1996 Oct 17;335(16):1169–75.
[3] Collins A. Clinical neurogenetics: Friedreich ataxia. Neurol Clin 2013 Nov;31(4): 1095–120.
[4] Delatycki MB, Williamson R, Forrest SM. Friedreich ataxia: an overview. J Med Genet 2000 Jan;37(1):1–8.
[5] Reetz K, et al. Biological and clinical characteristics of the European Friedreich's Ataxia Consortium for Translational Studies (EFACTS) cohort: a cross-sectional analysis of baseline data. Lancet Neurol 2016;14(Issue 2):174–82.
[6] Reetz K1, Dogan I1, Hohenfeld C1, Didszun C1, Giunti P1, Mariotti C1, Durr A1, Boesch S1, Klopstock T1, Rodríguez de Rivera Garrido FJ1, Schöls L1, Giordano I1, Bürk K1, Pandolfo M1, Schulz JB2, EFACTS Study Group. Nonataxia symptoms in Friedreich ataxia: report from the registry of the European friedreich's ataxia consortium for translational studies (EFACTS). Neurology 2018;91(10):e917–30. https://doi.org/10.1212/WNL.0000000000006121. 2018 Sep 4.
[7] Chamberlain S, Koenig M, Richter A, Palau F, Pandolfo M. Molecular analysis of the Friedreich's ataxia locus. Adv Neurol 1993;61:193–204.
[8] De Biase I, Chutake YK, Rindler PM, Bidichandani SI. Epigenetic silencing in Friedreich ataxiia is asssociated with depletion of CTCF (CCCTC-Binding factor) and antisense transcription. PLoS One 2009;4:e7914.
[9] Bürk K, Schulz SR, Schulz JB. Monitoring progression in Friedreich ataxia (FRDA): the use of clinical scales. J Neurochem 2013 Aug;126(Suppl 1):118–24.
[10] Wilson CL, Fahey MC, Corben LA, Collins VR, Churchyard AJ, Lamont PJ, et al. Quality of life in Friedreich ataxia: what clinical, social and demographic factors are important? Eur J Neurol 2007 Sep;14(9):1040–7.
[11] Weidemann F, Rummey C, Bijnens B, Störk S, Jasaityte R, Dhooge J, et al. The heart in Friedreich ataxia: definition of cardiomyopathy, disease severity, and correlation with neurological symptoms. Circulation 2012 Feb 29;125(13): 1626–34. https://doi.org/10.1161/CIRCULATIONAHA.111.059477. Epub 2012 Feb 29.
[12] Reetz, KathrinNachbauer Wolfgang, et al. Progression characteristics of the European Friedreich's ataxia consortium for translational studies (EFACTS): a 2 year cohort study. Lancet Neurol December 2016;15(Issue 13):1346–54. 1346 – 1354 Volume 15, No. 13.
[13] Witten IH, Frank E. Data mining: practical machine learning tools and techniques. second ed. Morgan Kaufmann; 2005. The 2nd edition of the data mining book.
[14] Cohen Jacob. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20(1):37–46. https://doi.org/10.1177/001316446002000104.
[15] Arora R, Suman. Comparative analysis of classification algorithms on different datasets using WEKA. Int J Comput Appl 2012;54:21–5. https://doi.org/10.5120/8626-2492.
[16] Quinlan JR. C4.5: Programs for machine learning. Morgan Kaufmann Publishers; 1993.
[17] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, WIttenm IH. The WEKA data mining software: an update. SIGKDD Explor 2009;11:10–8. https://doi.org/10.1145/1656274.1656278.
[18] David SK, Saeb ATM, Al Rubeaan K. Comparative analysis of data mining tools and classification techniques using WEKA in medical bioinformatics. Comput Eng Intell Syst 2013;4:28–38. https://doi.org/10.15373/22501991/JAN2013/8.
[19] Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286:531–7. https://doi.org/10.1126/science.286.5439.531.
[20] Salama GI, Abdelhalim MB, Zeid M. Experimental comparison of classifiers for Breast cancer diagnosis. In: 2012 7th International conference on computer engineering & systems, Cairo, 27-29 November 2012; 2012. p. 180–5.
[21] Al Jarullah AA. Decision tree discovery for the diagnosis of type II diabetes. In: Innovations in Information Technology (IIT), International Conference. 303–307; 2011. p. 25–7.
[22] Pousset F, Legrand L, Monin ML, Ewenczyk C, Charles P, Komajda M, Brice A, Pandolfo M, Isnard R, Tezenas du Montcel S, Durr A. JAMA A 22-year follow-up

study of long-term cardiac outcome and predictors of survival in Friedreich ataxia. Neurology 2015;72(11):1334–41. https://doi.org/10.1001/jamaneurol.2015.1855. Nov.

[23] Rosano GM, Panina G. Therapie 1999 May-Jun;54(3):381–5 [Oestrogens and the heart].

[24] Wong A, Yang J, Cavadini P, Gellera C, Lonnerdal B, Taroni F, Cortopassi G. The Friedreich's ataxia mutation confers cellular sensitivity to oxidant stress which is rescued by chelators of iron and calcium and inhibitors of apoptosis. Hum Mol Genet 1999;8:425–30 [PubMed].

[25] Richardson TE, Yu AE, Wen Y, Yang S-H, Simpkins JW. Estrogen prevents oxidative damage to the mitochondria in Friedreich's ataxia skin fibroblasts. PLoS One 2012; 7(4):e34600. https://doi.org/10.1371/journal.pone.0034600.

[26] Sandstede J, Lipke C, Beer M, et al. Eur Radiol 2000;10:438. https://doi.org/10.1007/s003300050072 [Age- and gender-specific differences in left and right ventricular cardiac function and mass determined by cine magnetic resonance imaging].

[27] Hudsmith Lucy E, Petersen Steffen E, Jane M, Francis, Robson Matthew D, Neubauer Stefan. Normal human left and right ventricular and left atrial dimensions using steady state free precession magnetic resonance imaging. J Cardiovasc Magn Reson 2005;7(5).

[28] Spirito P, Bellone P, Harris KM, Bernabo P, Bruzzi P, Maron BJ. Magnitude of left ventricular hypertrophy and risk of sudden death in hypertrophic cardiomyopathy. N Engl J Med 2000;342:1778–85. https://doi.org/10.1056/NEJM200006153422403.

[29] Bourke T, Keane D. Friedreich's Ataxia: a review from a cardiology perspective. Ir J Med Sci 2011;180:799. https://doi.org/10.1007/s11845-011-0744-y.and.refs.therein.

[30] Regner SR1, Lagedrost SJ, Plappert T, Paulsen EK, Friedman LS, Snyder ML, Perlman SL, Mathews KD, Wilmot GR, Schadt KA, Sutton MS, Lynch DR. Am J Cardiol 2012 Feb 1;109(3):401–5. https://doi.org/10.1016/j.amjcard.2011.09.025. Epub 2011 Nov 10. Analysis of echocardiograms in a large heterogeneous cohort of patients with friedreich ataxia.

[31] Peverill RE, Donelan L, Louise A, Corben LA, Delatycki MB. Differences in the determinants of right ventricular and regional left ventricular long-axis dysfunction in Friedreich ataxia. PLoS One 2018;13(12). 2018.

[32] Koeppen AH, Qian J, Travis AM, Sossei AB, Feustel PJ, Mazurkiewicz JE. Microvascular pathology in Friedreich cardiomyopathy. Histol Histopathol 2019: 18132. https://doi.org/10.14670/HH-18-132. 2019 Jun 5.

[33] Perdomini M, Hick A, Puccio H, Pook MA. Animal and cellular models of Friedreich ataxia. J Neurochem 2013;(126 Suppl 1):65–79. https://doi.org/10.1111/jnc.12219. Aug.

[34] Chung AK, Das SR, Leonard D, Peshock RM, Kazi F, Abdullah SM, Canham RM, Levine BD, Drazner MH. Women have higher left ventricular ejection fractions than men independent of differences in left ventricular volume. Circulation 2006;113: 1597–604.

[35] Hewer RL. Study of fatal cases of Friedreich's ataxia. Br Med J 1968;3(5619): 649–52.