HEART FAILURE RISK PREDICTION IDSS

Alice FAN CAN, Daniela FURLAN GONZALEZ, Mathilde POLIZZI, I-chun YEH

IDSS-Universitat Politècnica de Catalunya

June 12, 2018

Professors: Miquel Sànchez
Karina Gibert

Table of contents

I. Introduction

Heart failure is when the heart is unable to pump sufficiently to maintain blood flow to meet the body's needs. Signs and symptoms commonly include shortness of breath, excessive tiredness and leg swelling. It has been considered as one of the deadliest human diseases worldwide and the accurate prediction of this risk would be vital for heart failure prevention and treatment.

The architecture of the IDSS is described in Fig 1. First phase is the acquisition of the patient's data regarding heart failure diagnosis. Then the data is computed in the data mining phase using artificial neural networks and fuzzy-AHP methods. The results are then presented to the health professional thought the user interface to perform the decision making. Once a treatment is chosen, the patient is carefully followed by doctors who can perform new analysis after a few time in the actuation phase.

This paper aims to describe the implementation of the Intelligence Decision Support System described by O.W. Samuel [1] step by step, starting with the analysis of the initial dataset, the processing of the data by means of data mining techniques, ANN and Fuzzy-AHP and finally the results of the system through the user interface to help users such as health professionals to understand better the patient's situation and make decisions.

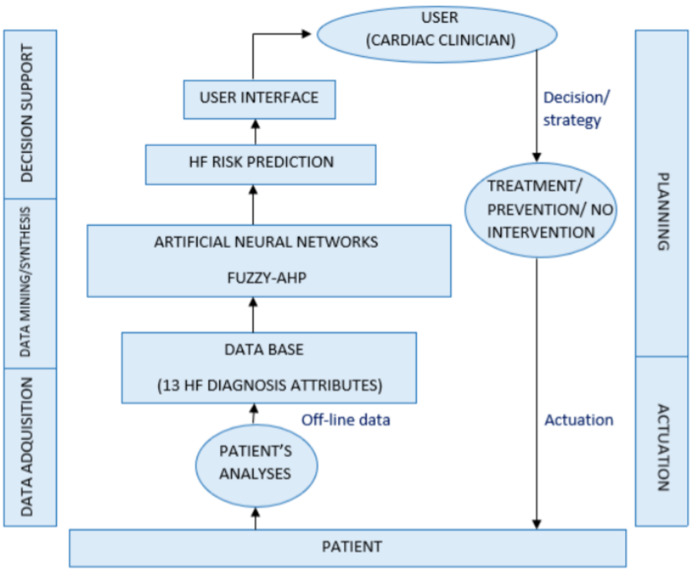


Fig 1. Functional architecture

II. IDSS for heart failure risk prediction

A. Data description and understanding

Heart disease dataset possesses 76 attributes but only 14 of them described in Table 1. are used for the analysis. The attributes represent the characteristics of the patient (age and sex) and the diagnosis of the heart failure prediction (chest pain, blood pressure, etc...). This dataset was taken from the UCI machine learning repository.

S/N	Attribute description	Attribute code	Alternatives	Alternative code	Range
1	Age (Years)	AGE	Young	YNG	<33
			Medium	MED	34–40
			Old	OLD	41–52
			Very old	VOLD	>52
2	Sex	SEX	Male	M	1
			Female	F	0
3	Chest pain type	CPT	Typical angina	TA	1
			Atypical angina	ATA	2
			Non-angina pain	NAP	3
			Asymptomatic	ASY	4
4	Resting blood pressure	RBP	Low	LOW	<128
			Medium	MED	128–142
			High	HIGH	143–154
			Very high	VHIGH	>154
5	Serum cholesterol	SCH	Low	LOW	<188
			Medium	MED	189–217
			High	HIGH	218–281
			Very high	VHIGH	>281
6	Fasting blood sugar	FBS	True	YES	1
			False	NO	0
7	Resting electrocardiographic results	RES	Normal	NOR	0
			ST-T abnormal	ST-AB	1
			Hypertrophy	HYPER	2
8	Maximum heart rate achieved	MHR	Low	LOW	<112
			Medium	MED	112–152
			High	HIGH	>152
9	Exercise induced angina	EIA	True	YES	1
			False	NO	0
10	Old peak	OPK	Low	LOW	<1.5
			Risk	RSK	1.5–2.55
			Terrible	TER	>2.55
11	Peak exercise slope	PES	Upsloping	UPS	1
			Flat	FLT	2
			Downsloping	DWS	3
12	Number of major vessels colored by fluoroscopy	VCA	Fluoroscopy-0	FL-0	0
			Fluoroscopy-1	FL-1	1
			Fluoroscopy-2	FL-2	2
			Fluoroscopy-3	FL-3	3
13	Thallium scan	THA	Normal	NOR	3
			Fixed defect	FDE	6
			Reversible defect	RDE	7

Table 1. Description of the 13 selected attributes from Heart Disease dataset

A quick understanding of the data was needed to implement the most suitable system for health professionals. On Fig 2 is shown a basic descriptive statistics analysis in order to better understand the dataset at hand. From the box plot, it can be seen that the dataset contains a few outliers for chest pain, resting blood pressure, serum cholesterol, and oldpeak attributes. Those are not a big issue for this study since it can be patient with unusual characteristics necessary to determine the heart risk presence. The distribution and histogram of the data help us have an overview about the how is distributed the values for each attributes. Finally the heatmap show there is no correlation between the attributes.

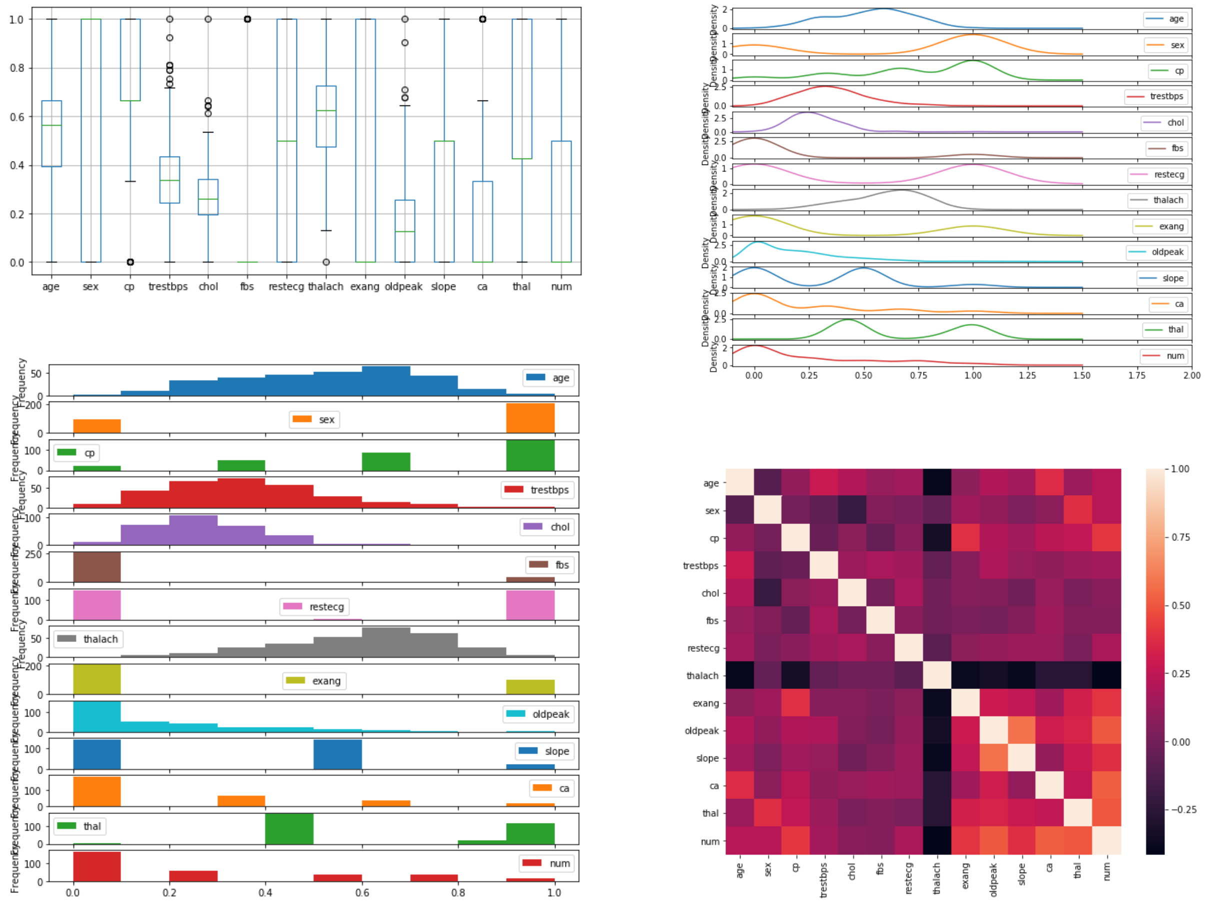


Fig 2. Data overview (boxplot, distribution, histogram and heatmap)

B. Flowchart of the IDSS model

The flowchart (Fig 3) starts with the patients' data saved in the database. In the preprocessing step, the data is normalized with a min-max function according to the values of each attribute. The scaling parameters are saved to normalize the new patients' data. Once the data is preprocessed, it is splitted in train and test data. Train data is used for training the hybrid model. The hybrid model consists of a fusion of fuzzy_AHP and ANN techniques. The results obtained are the predictions whether a patient has a risk of suffering a heart failure (HF). Finally, the prediction function of the hybrid model is used in the user interface to predict new patients risk of HF.

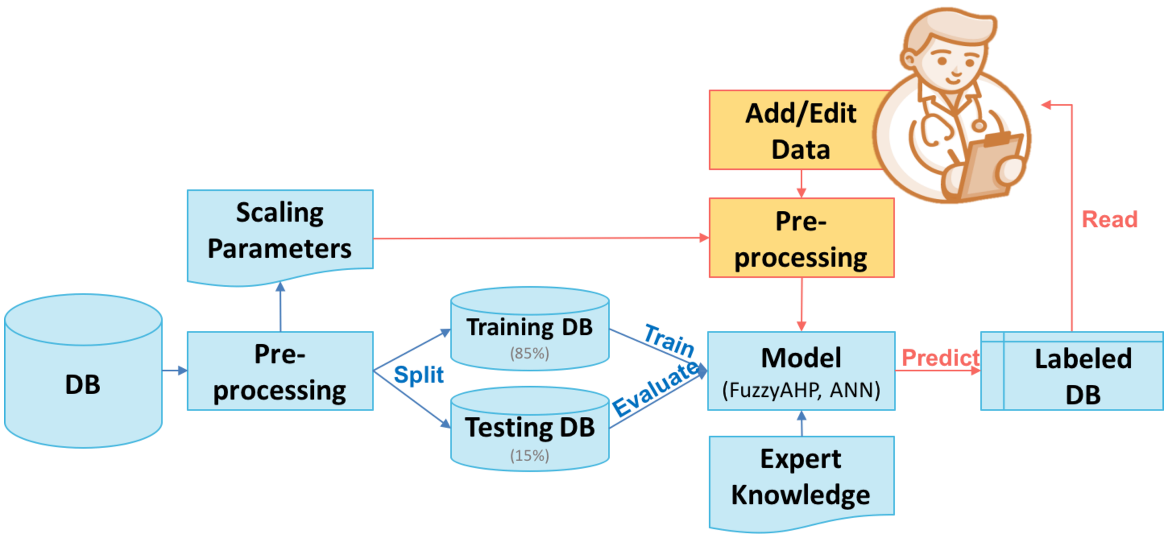


Fig 3. Flowchart of the IDSS model. Orange part is the user interface and blue part is the train/test model

C. Data pre-processing

Pre-processing is an important step in data mining to get the best quality of the dataset. Quality of data also means quality of results and decisions. In the preprocessing step it is relabelled the true labels of the patients. The 14th attribute in the database is recategorized, 0 values remain the same and for outputs ranged from 1 to 4 are relabelled as 1. Moreover, this one class label is transformed to two class label in order properly train the Artificial Neural Networks. Data is also normalized with min-max function [Eq.1], according to each attribute values and splitted into labels and features. Like in O.W. Samuel paper [1], the 65% of data is divided for train set, 20% for validation and 15% for test set. Patients with missing values are removed. No missing data treatment such as KNN is applied. More information about this decision is provided in the A)Experiments with different IDSS techniques section from III-Results part.

$$norm\_data = \frac{(x - minVal)}{(maxVal - minVal)} \text{ [Eq.1]}$$

D. IDSS techniques

Fuzzy - AHP technique

With this technique, it is obtained the initial weights of the ANN technique. To calculate these weights, first is done the hierarchical structure (Fig.4) of the attributes (variables) and its alternatives (different possible values of each variable). Once is structured the variables and its alternatives in a hierarchical structure, the second step is to obtain the weights of each attribute. To calculate the weights of each attribute first is done the comparison matrix (Fig.5) . Each value of the comparison matrix is a pairwise comparison of each attribute with respect to the other attributes. These evaluations ranges from 1 to 9, and normally are realized by experts. However, in our implementation these values were calculated from the weight attribute results showed in table 3 of the paper [1] (Fig.6). The weights results of the paper[1] were obtained by applying a Fuzzy_AHP technique and based on experts' knowledge. Our weights, based on paper weights, are obtained with AHP method. Our comparison matrix has a consistency ratio (CR) of 0.049. A comparison matrix is consistent if CR is less than the 10%, the result is 4.9%, so it could be concluded that the comparison matrix is consistent. The weights we have obtained, with AHP model from the Fuzzy_AHP results of the paper, are different from the ones showed in the paper (Fig 6). However, the ranking of the attributes is the same. Once the comparison matrix is obtained and it is consistent, the weights are calculated by calculating the normalized principal eigenvector.

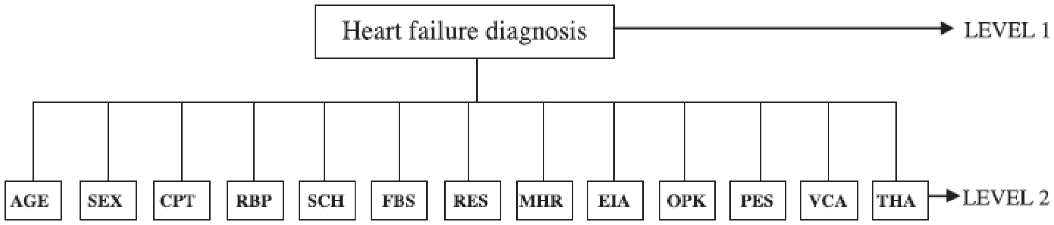


Fig 4. Hierarchical structure showing the goal (Level 1) and attributes (Level 2) . Retrieved from paper[1]

$$\begin{vmatrix}
1 & 6 & 1/2 & 2 & 3 & 3 & 4 & 1/2 & 2 & 1/2 & 5 & 1 & 1/4 \\
1/6 & 1 & 1/9 & 1/5 & 1/4 & 1/4 & 1/3 & 1/9 & 1/5 & 1/7 & 1/2 & 1/6 & 1/9 \\
4 & 9 & 1 & 5 & 6 & 6 & 7 & 2 & 5 & 3 & 8 & 4 & 2 \\
1/2 & 5 & 1/5 & 1 & 2 & 2 & 3 & 1/5 & 1 & 1/3 & 4 & 1/2 & 1/5 \\
1/3 & 4 & 1/6 & 1/2 & 1 & 1 & 2 & 1/6 & 1/2 & 1/4 & 3 & 1/3 & 1/6 \\
1/3 & 4 & 1/6 & 1/2 & 1 & 1 & 2 & 1/6 & 1/2 & 1/4 & 3 & 1/3 & 1/6 \\
1/4 & 3 & 1/7 & 1/3 & 1/2 & 1/2 & 1 & 1/7 & 1/3 & 1/5 & 2 & 1/4 & 1/7 \\
4 & 9 & 1/2 & 5 & 6 & 6 & 7 & 1 & 5 & 3 & 8 & 4 & 1 \\
1/2 & 5 & 1/5 & 1 & 2 & 2 & 3 & 1/5 & 1 & 1/3 & 4 & 1/2 & 1/5 \\
2 & 7 & 1/3 & 3 & 4 & 4 & 5 & 1/3 & 3 & 1 & 6 & 2 & 1/3 \\
1/5 & 2 & 1/8 & 1/4 & 1/3 & 1/3 & 1/2 & 1/8 & 1/4 & 1/6 & 1 & 1/5 & 1/8 \\
1 & 6 & 1/4 & 2 & 3 & 3 & 4 & 1/4 & 2 & 1/2 & 5 & 1 & 1/4 \\
4 & 9 & 1/2 & 5 & 6 & 6 & 7 & 1 & 5 & 3 & 8 & 4 & 1
\end{vmatrix}$$

Fig 5. Comparison matrix, obtained from paper weights

Attribute	Paper weights	Our weights
AGE	0.0822	0.0679
SEX	0.0287	0.0119
CPT	0.3333	0.2116
RBP	0.0645	0.0459
SCH	0.0559	0.031
FBS	0.0531	0.031
RES	0.0452	0.0216
MHR	0.1235	0.176
EIA	0.0696	0.0459
OPK	0.0997	0.0976
PES	0.0386	0.0157
VCA	0.0849	0.0679
THA	0.1208	0.176

Fig 6. Weights from the paper and calculated weights for each attribute

ANN technique

With Artificial Neural Network technique it is obtained the patient's predictions. Like in O.W. Samuel et al. [1] paper, it was build an ANN of 13 - 10 -2 (Fig.7). The input layer consists of 13 nodes corresponding of each one of the attributes. The hidden layer consists of a group of 10 nodes and the output layer is formed by two output nodes. The two values obtained are, first value is the percentage of absence of HF risk, and the second value is the percentage of presence of HF risk. The input values are first multiplied by fuzzy_AHP weights before fitting the ANN model.

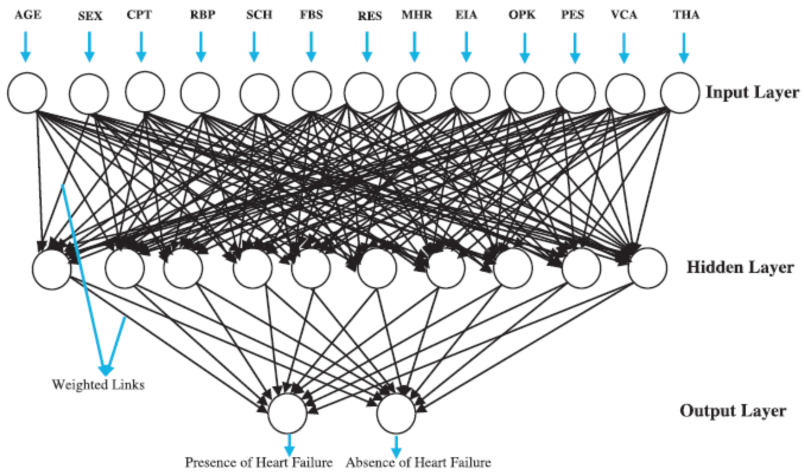
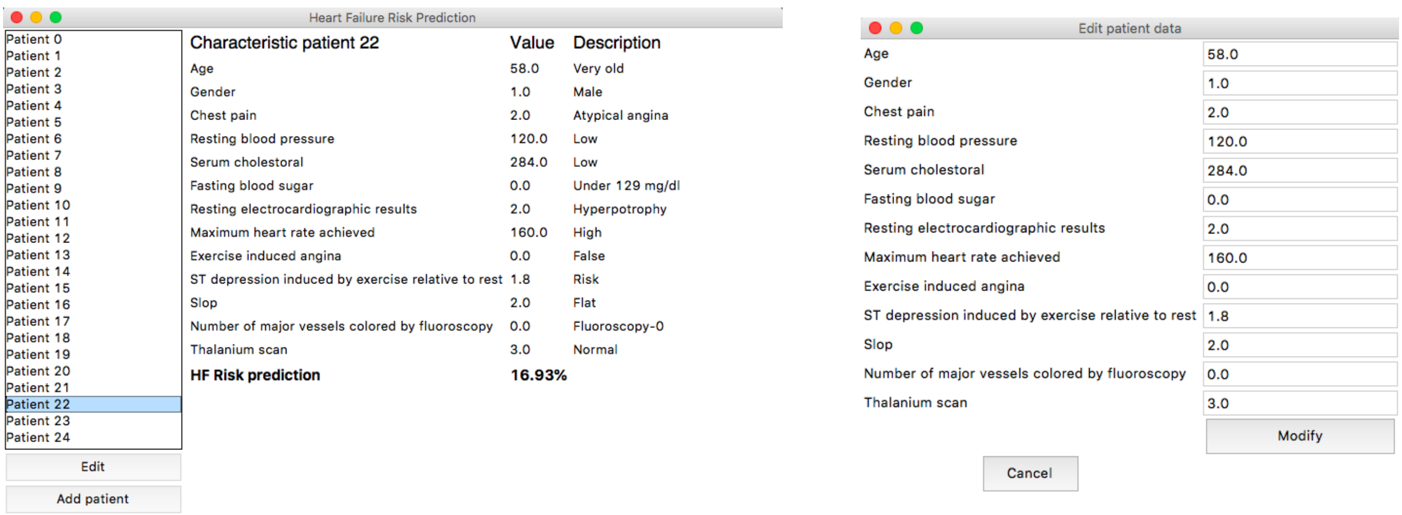


Fig 7. ANN model of 13-10-2 structure. Retrieved from [1]

E. Graphic user interface

The user interface aims to present results of the processing to health professional so they can perform the decision making. It has to be the more user-friendly and clear possible. A desktop app has been implemented and has the following features:

1. See the list of patients
2. Select patient's diagnosis
3. See the prediction, in other words the percentage of heart failure risk presence (results obtained from AHP-ANN model)
4. Edit patient's data
5. Add new patient data and compute prediction



(a) Patient diagnosis and prediction *(b) Edit patient data form*

Fig.8. Graphic User Interface

III. Results

A. Experiments with different IDSS techniques

It was done many experiments with different data treatments, techniques and structures combinations to see which combination provided better results. The results are shown in the table of Fig.8. Possible combinations are: Scaling method, which could be min-max or normalization methods. Missing data treatment, which could be new_category, MICE, KNN with one neighbour or KNN with three neighbours. Attribute weight: what means applying fuzzy_ahp technique or not. Fix attribute weight, if it is not fixed it means that the weights are learned from the ANN model, and the ANN architecture that could be 13-10-2 or 13-10-1, depending on the number of output nodes.

Scaling Method	*Missing data Treatment	Attribute Weight	Fix Attribute W	ANN	test acc (train acc)
min_max	replace_mean	Yes	Yes	13-10-2	80.43% (86.38%)
normalize	replace_mean	Yes	Yes	13-10-2	82.61% (86.38%)
min_max	replace_med	Yes	Yes	13-10-2	78.26% (85.60%)
normalize	replace_med	Yes	Yes	13-10-2	80.43% (85.60%)
min_max	MICE	Yes	Yes	13-10-2	80.43% (85.99%)
normalize	MICE	Yes	Yes	13-10-2	82.61% (86.38%)
min_max	knn_1	Yes	Yes	13-10-2	78.26% (85.99%)
normalize	knn_1	Yes	Yes	13-10-2	80.43% (86.38%)
min_max	knn_3	Yes	Yes	13-10-2	78.26% (85.99%)
normalize	knn_3	Yes	Yes	13-10-2	80.43% (86.38%)
min_max	x	Yes	Yes	13-10-2	95.56% (84.13%)
normalize	x	Yes	Yes	13-10-2	93.33% (84.92%)
min_max	x	No	Yes	13-10-2	91.11% (85.32%)
min_max	x	Yes	No	13-10-2	84.44% (84.92%)
min_max	x	Yes	Yes	13-10-1	93.33% (83.73%)

Fig 9. Comparison of different model approaches

It can be seen in the table that the combination that performs better is the one with the scaling method min-max, no missing data treatment, application of fuzzy-ahp, fix attribute weight and an ANN architecture of 13-10-2. This model is the one explained in this paper. The model used in O.W. Samuel [1] is the same but without scaling method.

B. Validation and evaluation of the selected model

Validation

For validating the hybrid model, Fuzzy-ahp and ANN, it was compared its performance with the performance of the conventional ANN method. Conventional ANN method consisted of the same ANN architecture but without multiplying fuzzy-AHP weights to the attributes. Validation was done using training and validating sets. The performance plots obtained are shown below. It could be seen that both models perform similar as both obtain similar loss function for train and validation sets. The main difference is the number of epochs, hybrid model needs 2000 epochs to stabilize and conventional model, after the epoch 1000, starts to overfit. Normally, the train error is lower than validation error. However, with this random partition of the sets, train error is greater than validation error. This could be due to data. In other words, in this partition, it could have been included in the training set many 'hard' cases to learn or the validation set could have mostly 'easy' cases to predict[2].

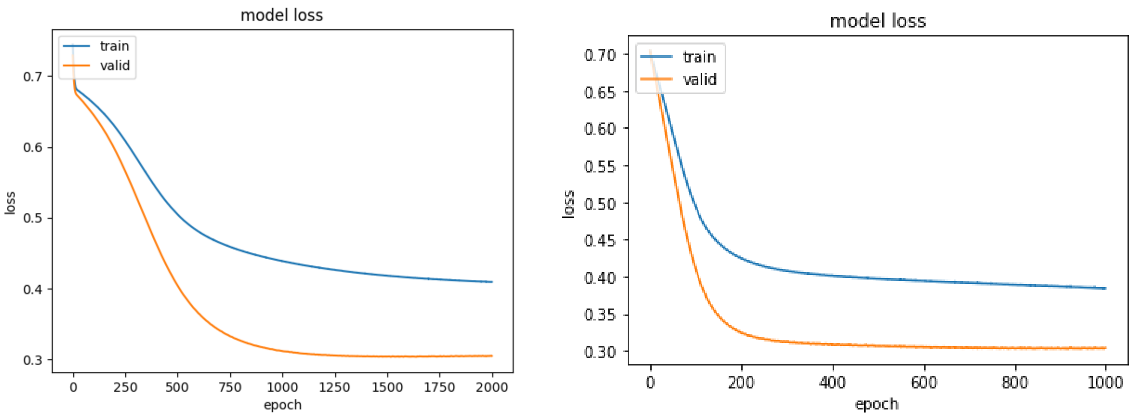


Fig 10. Performance plot of the hybrid model (left) and performance plot of the conventional ANN model (right)

Evaluation

For evaluating our model, it was also used the measures obtained from both models, hybrid method and conventional method. For the evaluation step it was used the test set. The results of test set are shown below in the confusion matrix plots(Fig 11). It can be noticed that hybrid method has no false positive label, contrary to conventional method. In hybrid method all HF risk absence cases are labelled correctly. In both models there are two cases of false negatives. Hybrid model does not improve false negatives cases.

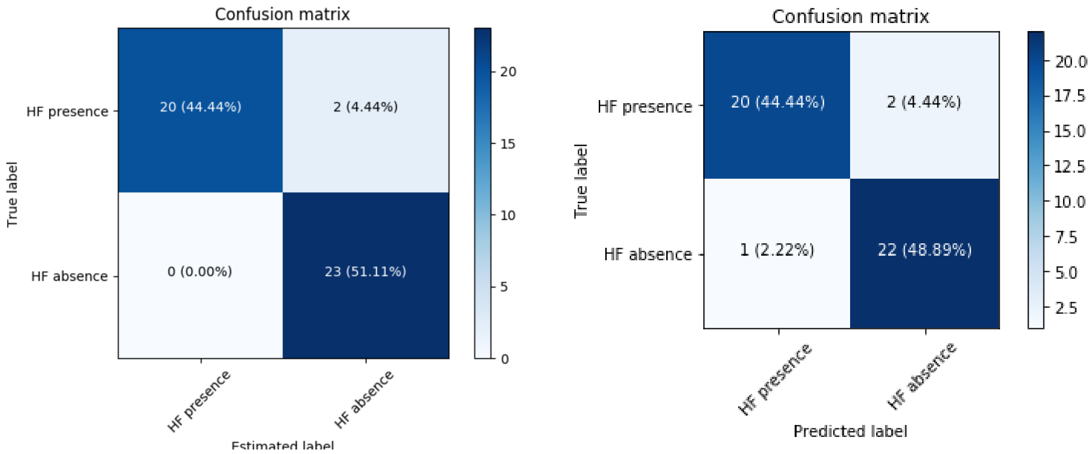


Fig 11. Confusion matrix of the test set with fuzzy_AHP and ANN model (left) and confusion matrix test set with conventional ANN technique (right)

For comparing the results of both methods it was also calculated indices such as the accuracy, f1 measure or specificity, among others. These indices are calculated from the results of the confusion matrices of Fig.11. In the table of Fig.12 it can be seen that hybrid model achieves a 95.56% of accuracy with test set compared to the 93.33% of accuracy of the test set with the conventional method. With test data it was also calculated F1 value, hybrid model achieved a 95.83% and conventional a 93.88%.

	hybrid method	conventional ANN method
Test Accuracy	95.56%	93.33%
Sensitivity	100.00%	95.65%
Specificity	90.91%	90.91%
Recall	100.0%	95.65%
Precision	92.00%	91.67%
F1	95.83%	93.62%

Fig 12. Indices obtained with hybrid and conventional ANN methods.

It was also plotted the ROC curves for model evaluation (Fig 13). It can be seen that the areas under ROC curves are very similar for both classes HF presence and HF absence.

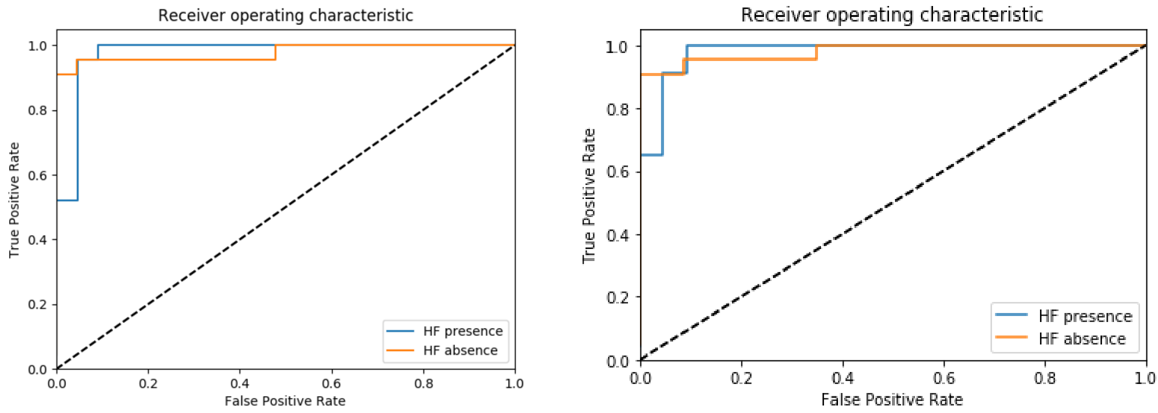


Fig 13. ROC curve from test data with hybrid model(left) and ROC curve of test data with conventional ANN technique (right).

IV. Conclusions

The obtained results show that, hybrid method obtained 2% higher accuracy than conventional method. We think this is not a big difference so as future work, we propose to study it in deep by using a bigger data set.

The IDSS implemented in this paper achieved better results, both for training and test sets, and for both models, hybrid and conventional techniques compared to O.W. Samuel[1]. The main differences detected between O.W. Samuel[1] implementation and ours are the number of epochs and the preprocessing step of scaling data, which is not explained in the paper.

The implemented IDSS includes the user interface modul which is not applied in the original paper. In addition, this module includes the function of predicting new patient's HF risk.

For this data it is obtained better results without missing data treatment. It could be due the fact that missing data treatment changes the original distribution.

V. Work plan

For developing this project Python version 3.5 and Github for sharing and versioning were used. As management tools, Trello and Google drive folder were used as well. Finally we used Slack for the meetings and to share instant thoughts.

With Gantt Diagram (Fig.14), work was encompassed by according delivering times for each part of the project. In the assignment grid (Fig.15) it is shown how were distributed the different tasks of the project.

A. Gantt Diagram

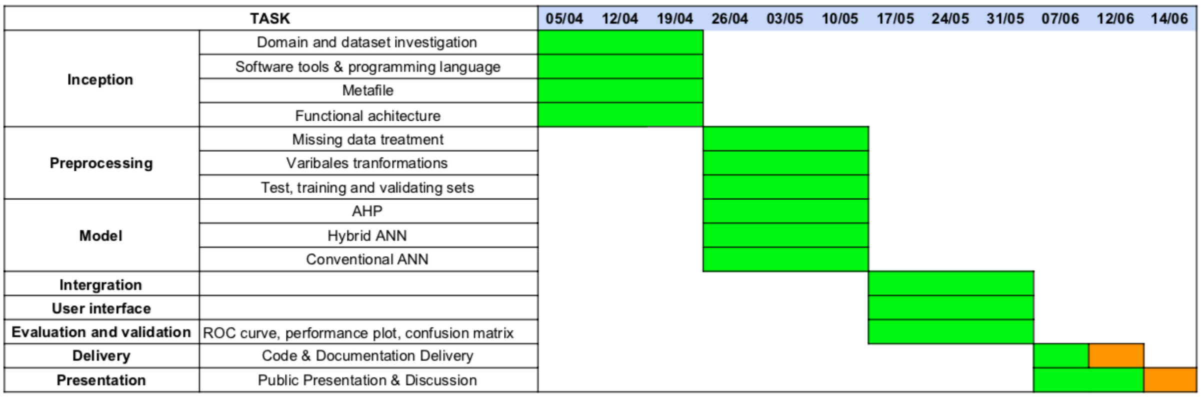


Fig 14. Gantt diagram

B. Assignment grid and Time sheet

TASK		Alice Fan		Daniela		Mathilde		I-Chun Yeh	
Inception	Domain and dataset investigation	✔		✔	4	✔	3	✔	3
	Software tools & programming language	✔		✔	0,1	✔	1	✔	1
	Functional achitecture			✔	1		0		0
Preprocessing	Missing data treatment	✔	5		1		0,5		0
	Varibales tranformations	✔			0		0	✔	0,3
	Test, training and validating sets			✔	1		0	✔	0,2
Model	AHP			✔	6		0		0
	Hybrid ANN				0		0	✔	5
	Conventional ANN				0		0	✔	3
Integration	Connect preprocessing, model & evalutation				0		0	✔	8
User interface	Implement GUI				0	✔	12		0
Evaluation & validation	ROC curve, performance plot, confusion matrix			✔	4		0		0
Report		✔		✔	5	✔	5	✔	8
Presentation	Public Presentation & Discussion	✔		✔	3	✔	2	✔	2
✔: Assigned task	**Total**		5		25,1		23,5		30,5

Fig 15. Assignment grid and Time sheet

V. References

[1] O.W. Samuel, G.M. Asogbon, A.K. Sangaiah, P. Fang, G. Li (2017) - *An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risK prediction*. Expert Syst. Appl. 68, 163-172.

[2]Stack Exchange. 2018. *Validation error less than training error?*. Available at: https://stats.stackexchange.com/questions/187335/validation-error-less-than-training-error. [Accessed 12 June 2018].