Diagnostic Feature Selection on Osteoporosis Clinical Data Using Genetic Algorithms

George C. Anastassopoulos[1], Adam Adamopoulos[1], Georgios Drosos[1], Konstantinos Kazakos[1] and Harris Papadopoulos[2,3]

[1]Democritus University of Thrace

[2]Frederick Research Center, Cyprus

[3]Frederick University, Cyprus

Outline

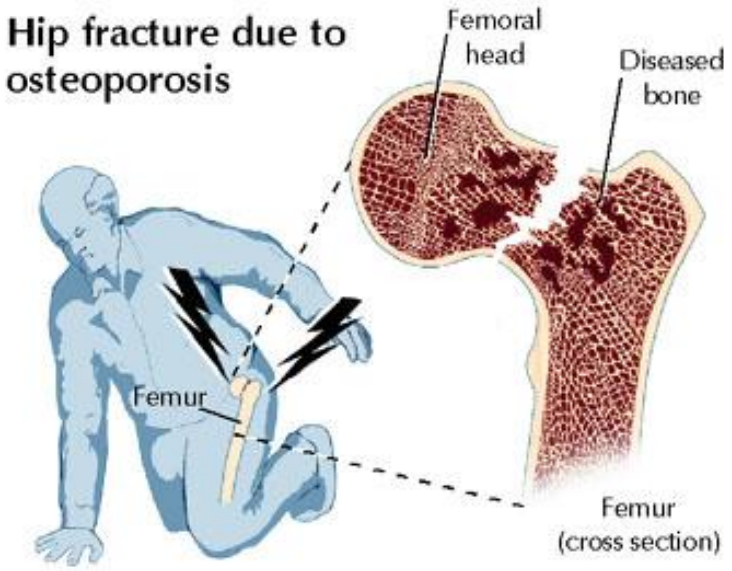
- Osteoporosis
- Overview and motivation of the project
  - **Development of New Venn Prediction Methods for Osteoporosis Risk Assessment**
- Why Venn prediction?
- Collected data
- Feature selection with a Genetic Algorithm
- Conclusions and future work

Osteoporosis

- Osteoporosis is a metabolic bone disease in which the bones become more and more fragile, leading to an increased risk of fracture.
- In the European Union one person breaks a bone because of osteoporosis every fifteen seconds.
- Often the first apparent symptom of osteoporosis is a broken bone, which is why the condition is also known as "the silent crippler".
  - Early detection and treatment of osteoporosis can decrease the fracture risk of a person to a minimum.

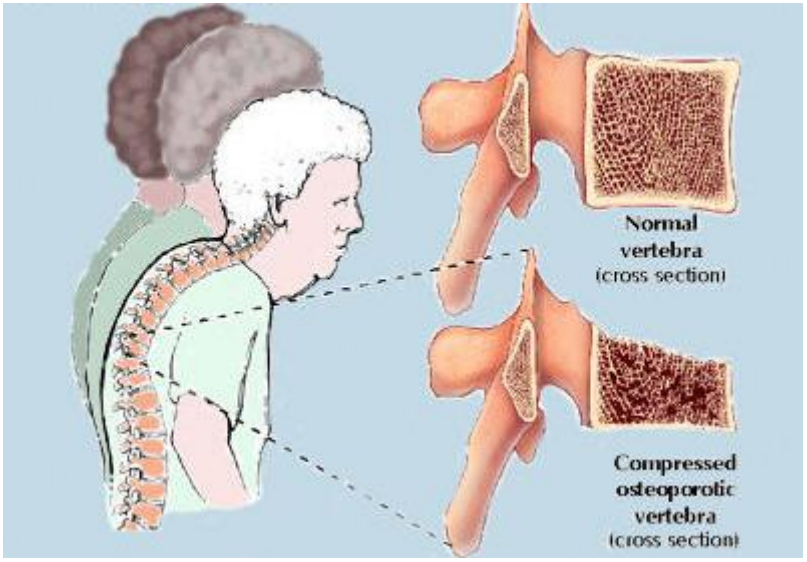
Osteoporosis

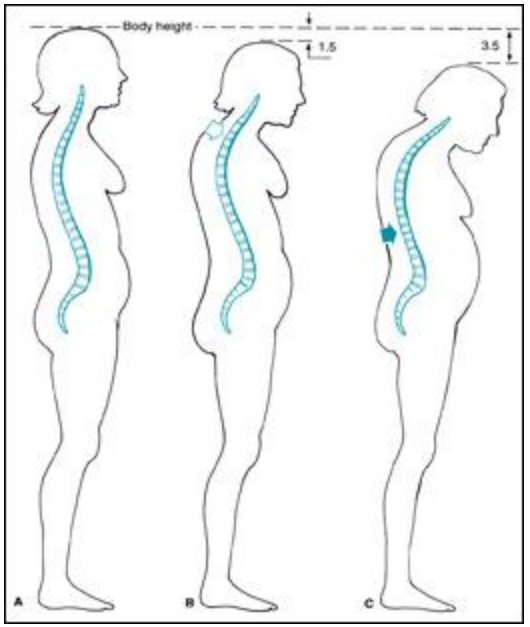
- Osteoporosis is a metabolic bone disease in which the bones become more and more fragile, leading to an increased risk of fracture.
- In the European Union one person breaks a bone because of osteoporosis every fifteen seconds.



Osteoporosis

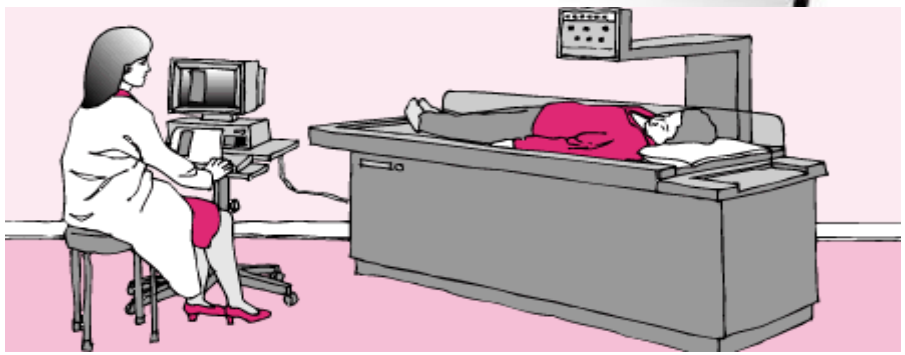
- Often the first apparent symptom of osteoporosis is a broken bone, which is why the condition is also known as "the silent crippler".
  - Early detection and treatment of osteoporosis can decrease the fracture risk of a person to a minimum.





Osteoporosis Detection

- Today's clinical routine is based on:
  - Dual Energy X-Ray Absorptiometry (DEXA) scan



T-score Value	Coding
≤ - 2.5	3 (Osteoporosis)
-2.5 – -1.5	2 (Osteopenia)
-1.5 – 0	1 (poor osteopenia)
≥ 0	0 (normal)

Overview of the Project

- Title: **Development of New Venn Prediction Methods for Osteoporosis Risk Assessment**
- Funded by the **European Regional Development Fund and the Cyprus Government through the Cyprus Research Promotion Foundation**
- Total budget: €163,144
- Partners

FREDERICK RESEARCH CENTRE

Royal Holloway University of London

ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ

CYPRUS SOCIETY AGAINST OSTEOPOROSIS AND MYOSKELETAL DISEASES

Overview of the Project

- Main objectives:
  - Development of new Venn Prediction techniques based on **Neural Networks** and **Support Vector Machines**
  - The application of these techniques to the problem of osteoporosis risk assessment
  - Development of a **risk assessment tool** for helping clinicians identify people that should undergo further testing

Motivation

- Classical Machine Learning techniques do not provide any indication about the likelihood of their predictions
- Therefore, most existing Medical Decision Support systems just output the most likely diagnosis of a new patient, without any other information
- This is a major drawback in a medical setting
  - A physician would need to know how reliable each output is
- The aim of this project is to overcome this problem with the use of a novel framework, called **Venn Prediction**, that provides **Probability Intervals**, which are **guaranteed** to contain **well-calibrated probabilities**

Motivation

- Venn Prediction will be applied to the problem of **assessing osteoporosis risk**
  - Helping clinicians identify people that should undergo further testing (DEXA scan)
- The resulting decision support tool will provide probabilistic intervals that will be much more informative than a plain yes or no output

Project Outline

- Data collection
  - Questionnaire containing osteoporosis risk factors
  - The data can be used in the future for the assessment of fracture risk
- Algorithm development
  - Development of new Venn Predictors
- Analysis of the collected data with AI techniques
  - Identification of the most important risk factors
- Algorithm application to the collected data
- Development of risk assessment system
- Pilot application and evaluation of the system

Experimental Demonstration of Venn Prediction (with Neural Networks)

- Two medical datasets
  - *Mammographic Mass*: discrimination between benign and malignant mammographic masses. 961 cases: 516 benign and 445 malignant.
  - *Pima Indians Diabetes*: forecasting the onset of diabetes mellitus in a high-risk population of Pima Indians. 768 cases: 500 positive and 268 negative.
- The NNs were trained with the scaled conjugate gradient algorithm minimizing cross-entropy error
  - Their outputs can be interpreted as probabilities
- Comparison between original NN and NN-VP

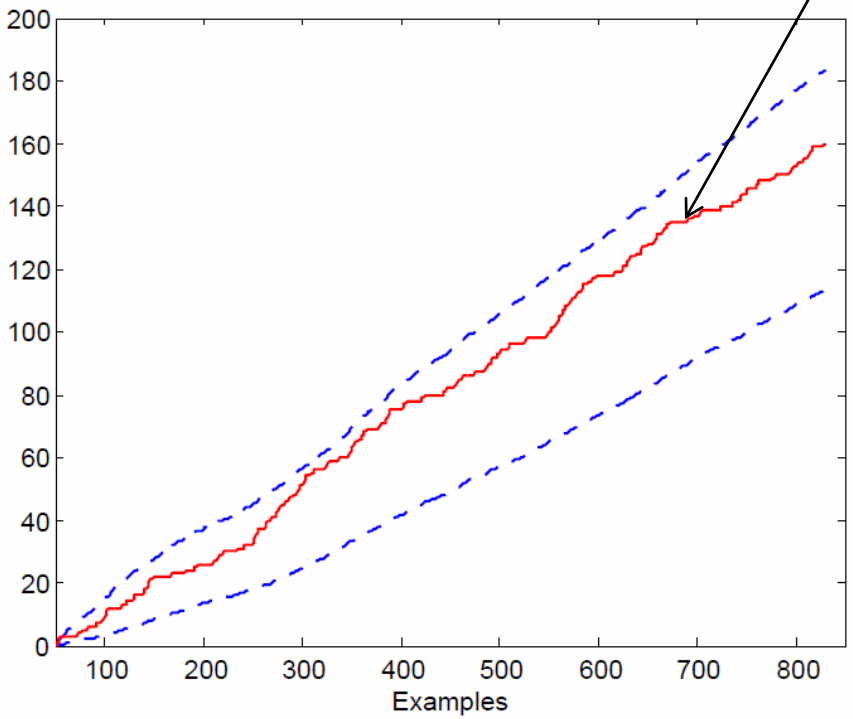
On-line Experiments

- Start with an initial training set containing 50 examples
- Predict each subsequent example
- After prediction each new example is added to the training set (with its true classification) for predicting the next examples

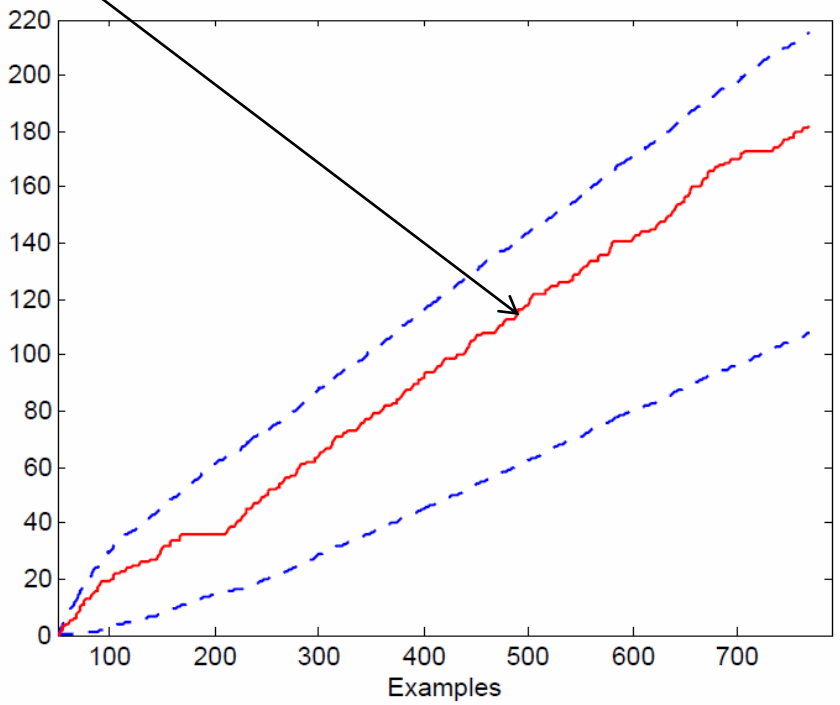
On-line Experiments: NN-VP

Cumulative Error Curve

$$E_n = \sum_{i=1}^{n} err_i$$



(a) Mammographic Mass

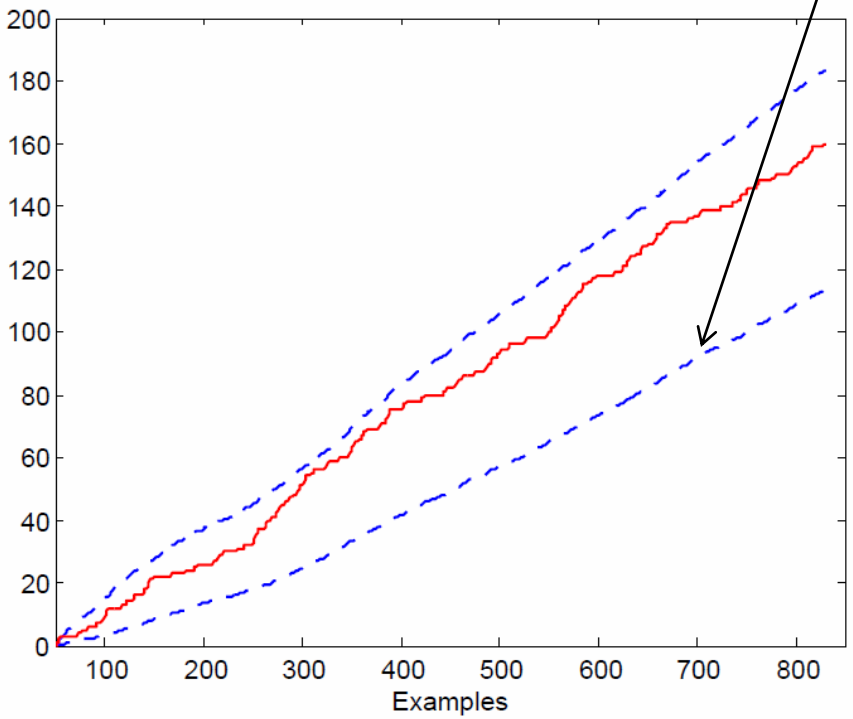


(b) Pima Indians Diabetes

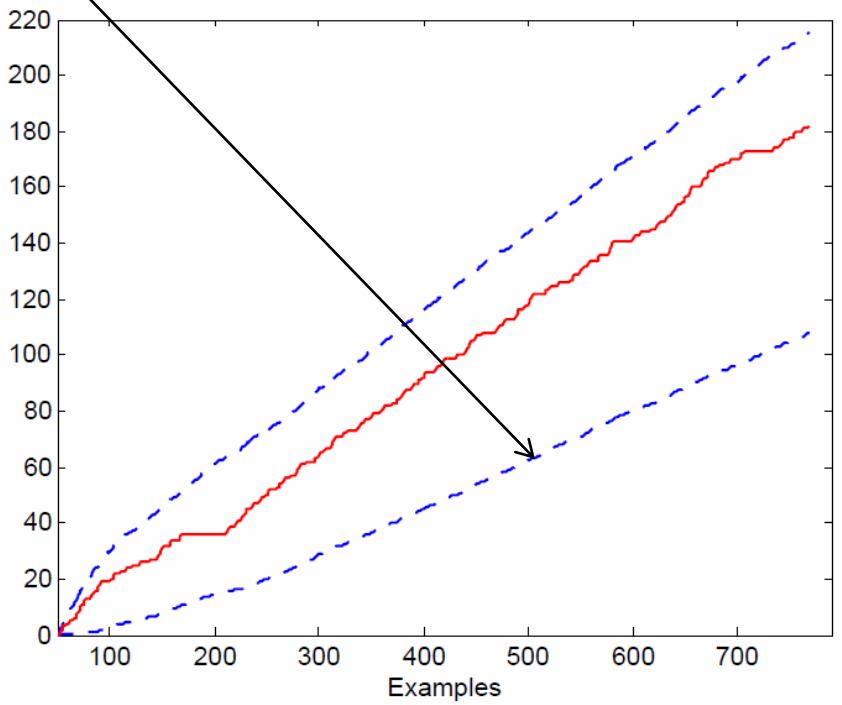
On-line Experiments: NN-VP

Cumulative Lower Error Probability Curve

$$LEP_n = \sum_{i=1}^{n} 1 - U(\hat{y}_i)$$



(a) Mammographic Mass

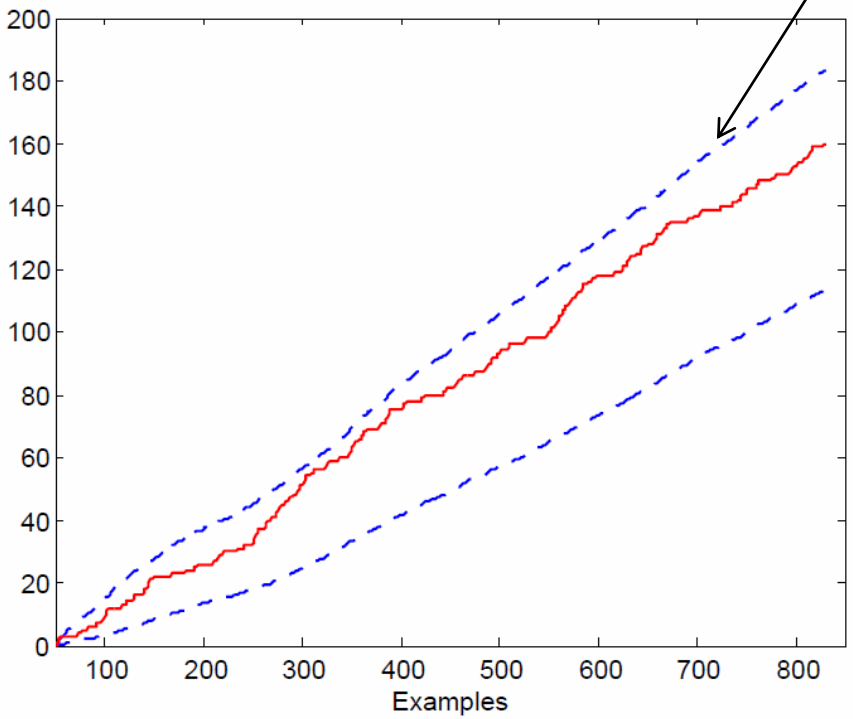


(b) Pima Indians Diabetes

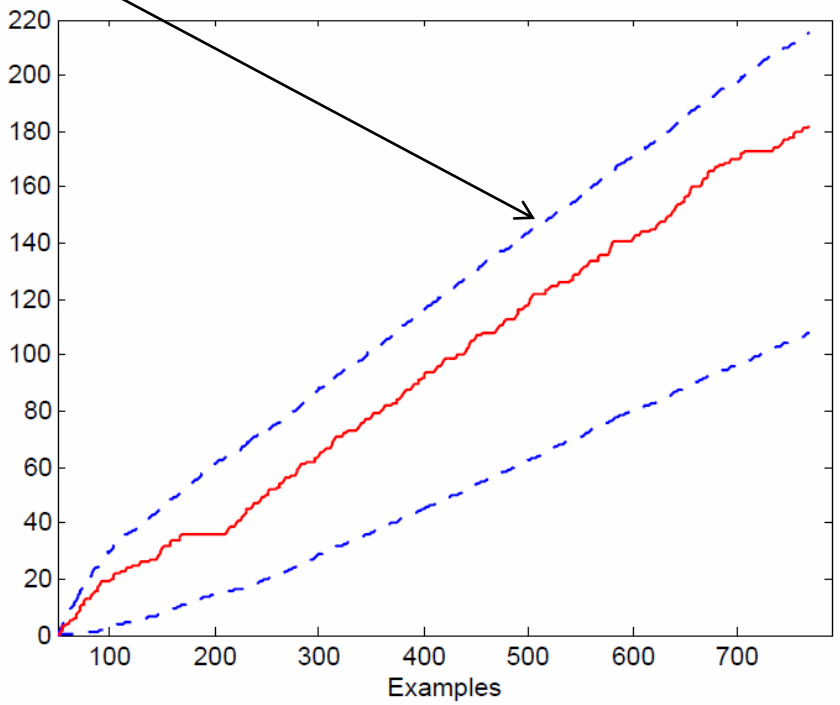
On-line Experiments: NN-VP

Cumulative Upper Error Probability Curve

$$UEP_n = \sum_{i=1}^{n} 1 - L(\hat{y}_i)$$



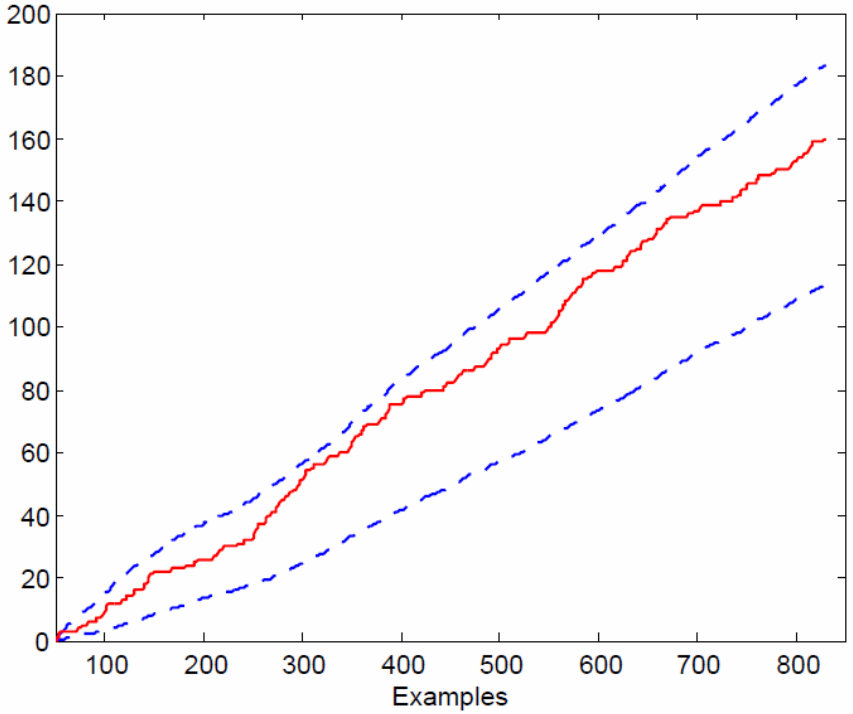
(a) Mammographic Mass



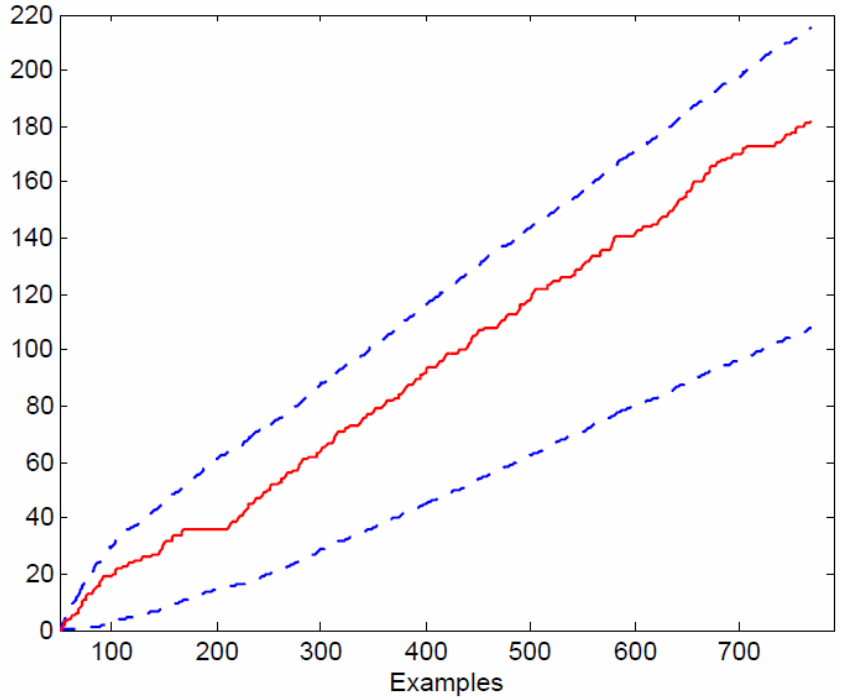
(b) Pima Indians Diabetes

On-line Experiments: NN-VP

- Both plots confirm that the probability intervals are well-calibrated



(a) Mammographic Mass

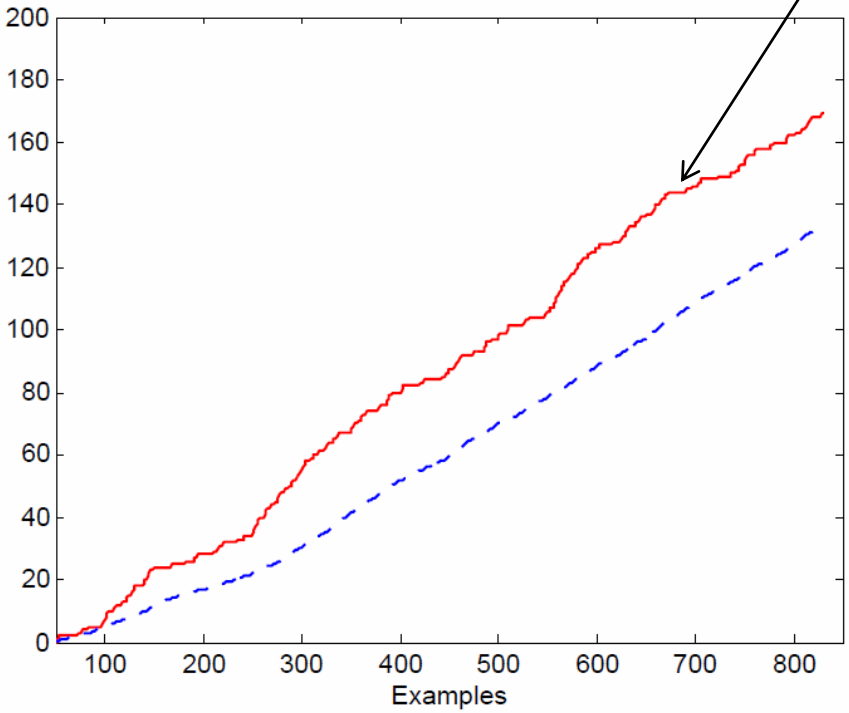


(b) Pima Indians Diabetes

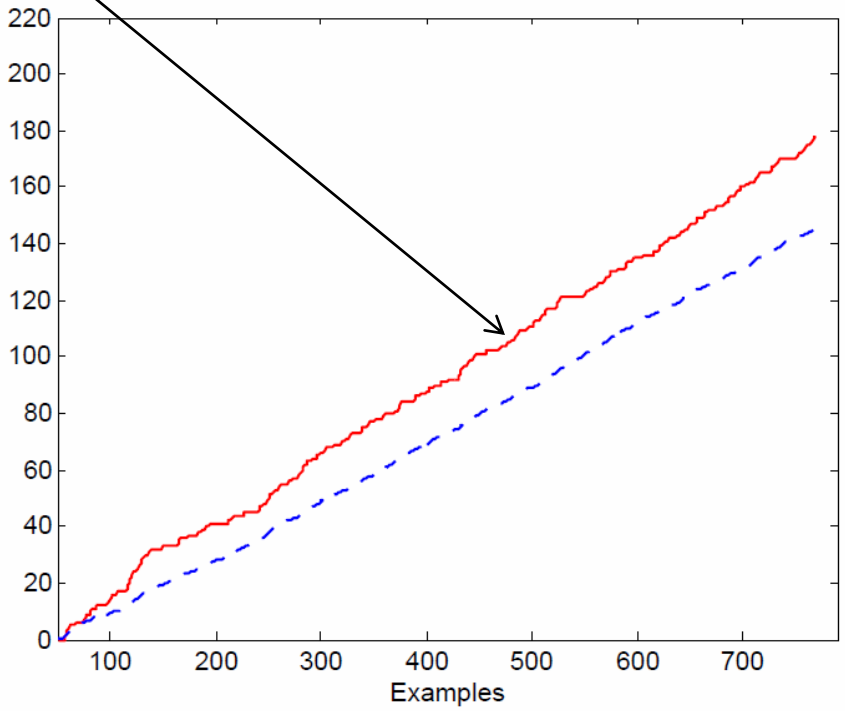
On-line Experiments: Conventional NN

Cumulative Error Curve

$$E_n = \sum_{i=1}^{n} err_i$$



(a) Mammographic Mass

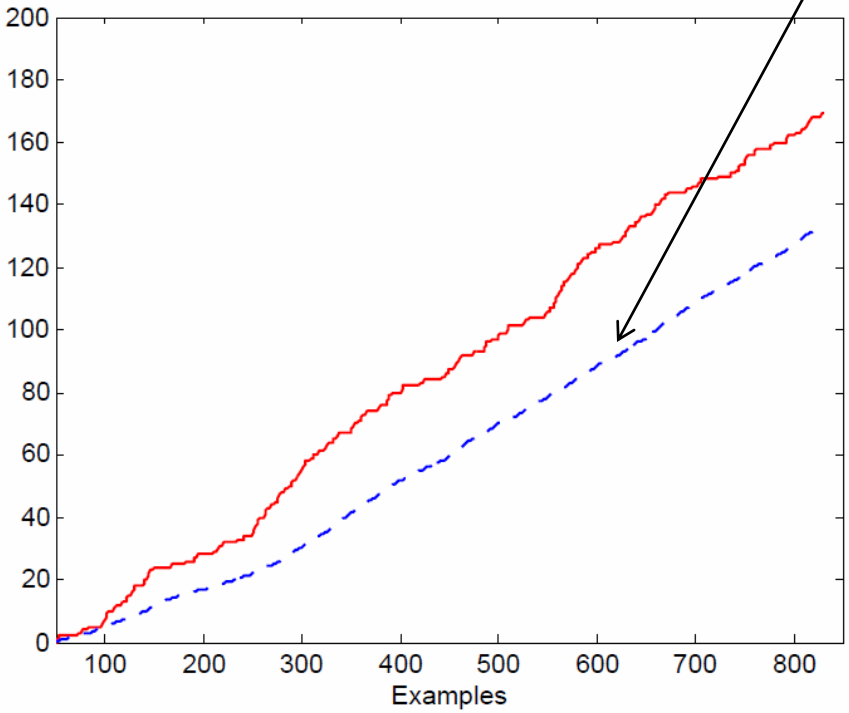


(b) Pima Indians Diabetes

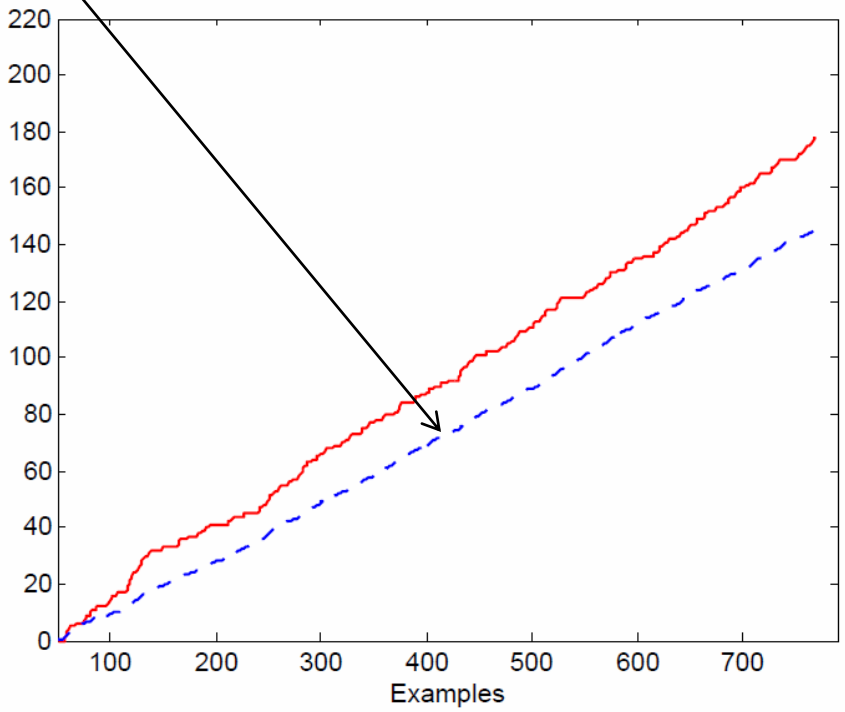
On-line Experiments: Conventional NN

Cumulative Error Probability Curve

$$EP_n = \sum_{i=1}^{n} |\hat{y}_i - \hat{p}_i|$$



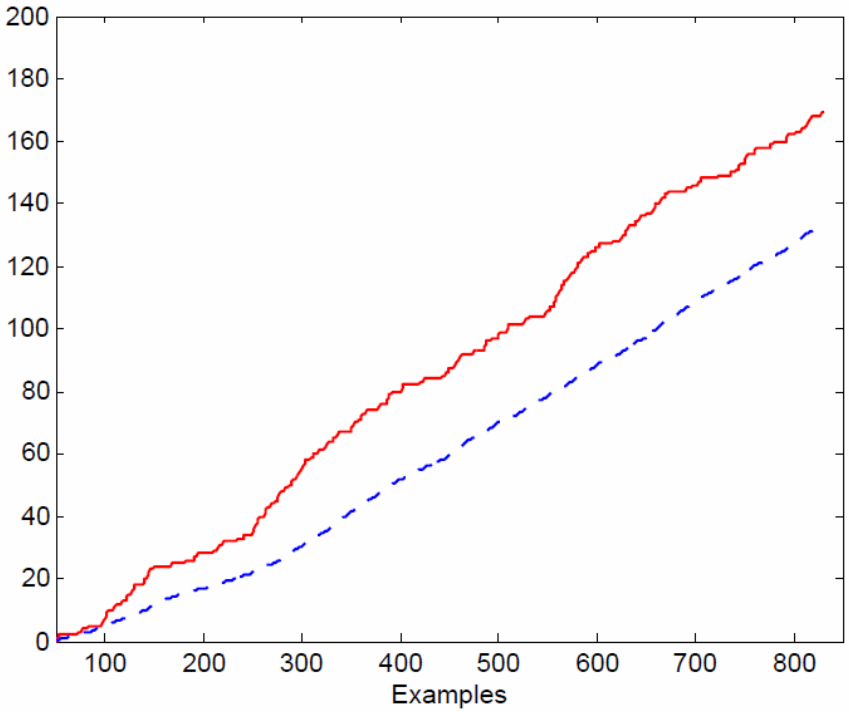
(a) Mammographic Mass



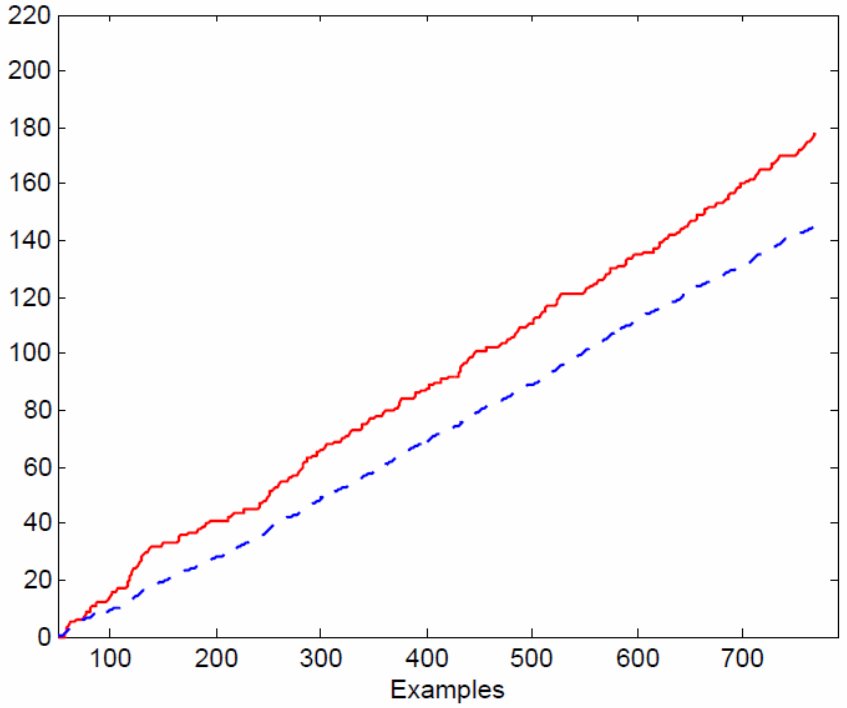
(b) Pima Indians Diabetes

On-line Experiments: Conventional NN

- The p-values of obtaining the resulting total number of errors given the probabilities produced by the ANN were 0.000179 and 0.000548 respectively (extremely unlikely!)



(a) Mammographic Mass



(b) Pima Indians Diabetes

Data (Standalone Information 1)

- Age
- Weight
- Height
- Sex
  - Male
  - Female
- Menstruation
  - Starting Age
  - Ending Age
- Number of pregnancies
- Smoking
  - Now
    - Years of Smoking
    - Cigarettes per day
  - In the past
    - Years of Smoking
    - Cigarettes per day
    - Never
- Alcohol (units/day)
- Caffeine (coffees per day)
- History of previous fracture
  - Yes/No
  - When
- Area of previous fracture
  - Hip
  - Spine
  - Wrist
- Type of fracture
  - High Energy
  - Low Energy
- Exercise more than 30 minutes, 3 times a week
- History of osteoporosis or fracture of the hip in parent

Data (Standalone Information 2)

- Family history of osteoporosis or another bone condition
- Loss of height more than 3 cm
- Kyphosis
- End of menstrual bleeding for more than 12 months for any reason except pregnancy
- Suffer from rheumatoid arthritis
- History of secondary osteoporosis
- Breastfeeding
- Avoid milk and other dairy products
- Men: Incompetence or lack of sexual desire or other symptoms associated with low testosterone levels
- Diarrhea due to chronic bowel disease
  - Yes
    - Condition
  - No
- Receive cortisone
  - Yes
    - Dosage
    - Reason
  - No
- Taking thyroxin
  - Yes
    - For how long
    - Reason
  - No
- Receive estrogen

Data (Previous conditions or surgeries)

- Neurogenic anorexia
- Malabsorption syndrome
- Chronic liver diseases
- Inflammatory bowel diseases
- Transplantation
- Chronic renal failure
- Prolonged immobilization
- Cushing's syndrome
- Epilepsy
- Insulin Dependent P. D
- Ovariectomy before menopause
- Chronic gastrointestinal disorders
- Disease of Paget's Disease
- Hyperthyroidism
- Parathyroid gland disease
  - Yes
    - Which
  - No

Data (Previous/current medicine or treatments)

- Steroids (prednisone, cortisone, etc.)
- Thyroxin
- Anticonvulsants (for seizures, epilepsy)
- Diuretics (Lasix)
- Heparin
- Chemotherapy

- Earlier treatment of osteoporosis (Yes/No)
  - Duration
  - Type of treatment
    - Alendronati 'Fosamax' or 'Fosavance'
    - Risedronati 'Actonel'
    - Zoledronati 'Aclasta' or 'Zomeda'
    - Raloxifeni 'Evista'
    - Strontio 'Protelos'
    - Parathormoni 'Forsteo'
    - Denosoymapi 'Prolia'
    - Kalsitonini 'Miacalcic'
    - Calcium + Vitamin D
    - Calcium
  - Dosage per day

Data (Bone density measurements)

- Lumbar Vertebrae
  - T-Score
  - Z-Score
- Hip
  - T-Score
  - Z-Score
- Neck
  - T-Score
  - Z-Score

Feature Selection with GA and NN

- A Genetic Algorithm together with a Neural Network approach was used to select the best feature subset (out of 33 features)
- Data collected at the Orthopedic Surgery Department of Alexandroupolis University Hospital
  - 589 cases
  - Random division: 80% for training and 20% for testing
- The GA chromosome encoded the inputs and the number of hidden units of the NN
- Fitness function:

$$f = \frac{MSE_p}{< MSE_t >} + \frac{I}{N} + \frac{H}{64}$$

Results

#Exp.	#Gen.	MSEp	I	Features	H
1	1	$1.63 \cdot 10^{-4}$	15	1 3 7 8 10 11 13 15 17 20 23 24 28 30 32	2
1	17	$1.43 \cdot 10^{-4}$	5	2 6 7 10 25	2
1	29	$1.37 \cdot 10^{-4}$	3	3 7 10	1
1	49	$1.37 \cdot 10^{-6}$	2	7 10	1
2	1	$1.85 \cdot 10^{-6}$	12	3 7 8 10 14 15 16 18 23 25 28 32	6
2	19	$1.65 \cdot 10^{-6}$	2	7 10	1

Conclusion

- Our final objective is to develop a medical decision support tool for assessing the probability of a person having osteoporosis.
- The Venn prediction framework, which is being used, provides probabilistic intervals that are guaranteed to contain well calibrated probabilities up to statistical fluctuations
  - Thus providing reliable decision support, which is of paramount importance in a medical setting
- In the particular work a GA was used in conjunction with NN for finding the best feature subset
- Future work
  - Further analysis with other feature subset selection techniques
  - Extension of the dataset
  - Analysis of the findings with doctors

Acknowledgments

- This work was supported by the European Regional Development Fund and the Cyprus Government through the Cyprus Research Promotion Foundation "ΔΕΣΜΗ 2009-2010" research contract ΤΠΕ/OPIZO/0609(BIE)/24 ("Development of New Venn Prediction Methods for Osteoporosis Risk Assessment").

Thank you for listening