**Heart Disease Prediction**

**Introduction**

Heart disease is a leading cause of morbidity and mortality worldwide .As failure to recognise atypical representations of such serious illness may lead to adverse outcomes, accurate diagnosis is crucial to ensure that patients are placed on the proper treatment pathway. Electronic medical records (EMR) can be used to improve the diagnosis ability along with measuring the quality of care. The rapid adoption of EMRs along with the necessity to enhance the quality of health care has incentivised the development of predictive modeling in the medical domain. An abundant amount of clinical information used for medical investigation is organised in unstructured narrative form, which is suitable for expressing medical concepts or events but challenging for analysis and decision support as gaining a full aspect of a patients medical history by reading through EMRs is significantly timeconsuming, especially when only a specific piece of information is needed. The difficulty of this process increases in the case of heart disease due to its complex progression, which regularly involves various factors including lifestyle and social factors as well as specific medical conditions. Various methods have been proposed in the field of clinical concept extraction, ranging from simple pattern matching to systems based on symbolic or statistical data and machine learning. Those previously proposed approaches have shown promising results but it is very difficult to reach that point due to the assiduous process of defining rules and extracting features. This is where deep learning comes in as this intriguing re-emerged concept can alleviate heavily human dependent efforts required for knowledge-based approaches and the lack of the ability of many conventional machine learning algorithms to learn without the necessity of careful feature engineering with considerable domain expertise.

**Objectives**

Automatic identification of heart disease risk factors in clinical narratives can expedite disease progression modelling and support clinical decisions. Existing practical solutions for cardiovascular risk detection are mostly hybrid systems entailing the integration of knowledge-driven and data-driven methods, relying on dictionaries, rules and machine learning methods that require a substantial amount of human effort. This project proposes the applicability of deep learning, a re-emerged data-driven technique, in the context of clinical text classification. Various deep learning architectures were devised and evaluated for extracting heart disease risk factors from clinical documents

**Problem Statement**

Early detection of heart failure Onset of HF is associated with a high level of disability, health care costs, and mortality (roughly 50% risk of mortality within 5 years of diagnosis). There has been relatively little progress in slowing the progression of this disease, largely because it is difficult to detect before actual diagnosis. As a consequence, intervention has primarily been confined to the time period after diagnosis, with little or no impact on disease progression. Earlier detection of HF could lead to improved outcomes through patient engagement and more assertive treatment with angiotensin-converting enzyme inhibitors or angiotensin II receptor blockers, mild exercise, reduced salt intake, and possibly other options. Previous work on early detection of HF has relied on conventional modeling techniques, such as logistic regression or support vector machine (SVM), that use features representing the aggregation of events in an observation window and exclude temporal relations among events in the observation window. In contrast, recurrent neural network (RNN) methods capture temporal patterns present in longitudinal data. RNN models have proven effective in many difficult machine learning tasks, such as image captioning and language translation. Extending these methods to health data is sensible.

**Review of Literature**

Applications of deep learning

Deep learning methods have recently led to a renaissance of neural network–based models. Pioneering studies introduced stacked restricted Boltzmann machines4 and stacked autoencoders,5 which showed impressive performance in image processing, employing the layer-wise pretraining technique. Since then, variations of neural network application have explored deep architectures in computer vision,6–8 audio processing,9,10 and natural language processing (NLP),3,11–13 among other fields. RNN models are naturally suited to temporal sequenced data, and several variants have been developed for sequenced features. Hochreiter and Schmidhuber14 proposed long short-term memory (LSTM), exhibiting impressive performance in numerous sequencebased tasks such as handwriting recognition,15 acoustic modeling of speech,16 language modeling,17 and language translation.18 Cho et al.3 proposed the gated recurrent unit (GRU) model, structurally similar to but simpler than LSTM, and showed comparable, if not better, performance.19 In the RNN work described herein, we used the GRU structure to model the temporal relations among health data from patient EHRs to predict the future diagnosis of HF. Health care applications of deep learning Researchers have recently started to apply deep learning methods to clinical applications. Lasko et al.20 used autoencoders to learn phenotypic patterns from serum uric acid measurements. Che et al.21 used deep neural networks with incremental learning on clinical time series data to discover physiologic patterns associated with known clinical phenotypes. Both works,20,21 however, focused on learning patterns from clinical records rather than predicting acinical event. Hammerla et al.22 applied restricted Boltzmann machines on time series data collected from wearable sensors to predict the disease state of Parkinson’s disease patients. Lipton et al.23 used LSTM for multilabel diagnosis prediction using pediatric ICU time series data (eg, heart rate, blood pressure, glucose level, etc.). Both of these studies22,23 used multivariate time series data from patients, which focused on very different clinical conditions, with continuous time series data. Our study focuses on early detection of HF for the general patient population based on widely available EHR data such as time-stamped codes (diagnosis, medication, procedure). Deep learning techniques have been recently applied to clinical text data (eg, PubMed abstracts, progress notes) using Skipgram12,24,25 to learn relationships among clinical processes or unified medical language system (UMLS) concepts. Choi et al.26 applied Skip-gram to longitudinal EHR data to learn low-dimensional representations for medical concepts such as diagnosis codes, medication codes, and procedure codes,27 and to learn representations of medical concepts. We borrowed from this prior work to leverage similar representation of medical concepts through Skip-gram but focus on temporal modeling using RNN for predicting HF.

**Data Collection**

>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

This describes the contents of the heart-disease directory.

This directory contains 4 databases concerning heart disease diagnosis.

All attributes are numeric-valued. The data was collected from the

four following locations:

1. Cleveland Clinic Foundation (cleveland.data)

2. Hungarian Institute of Cardiology, Budapest (hungarian.data)

3. V.A. Medical Center, Long Beach, CA (long-beach-va.data)

4. University Hospital, Zurich, Switzerland (switzerland.data)

Each database has the same instance format. While the databases have 76

raw attributes, only 6 of them are actually used. Thus I've taken the

liberty of making 2 copies of each database: one with all the attributes

and 1 with the 6 attributes actually used in past experiments.

The authors of the databases have requested:

...that any publications resulting from the use of the data include the

names of the principal investigator responsible for the data collection

at each institution. They would be:

1. Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D.

2. University Hospital, Zurich, Switzerland: William Steinbrunn, M.D.

3. University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.

4. V.A. Medical Center, Long Beach and Cleveland Clinic Foundation:

Robert Detrano, M.D., Ph.D.

>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

1. Title: Heart Disease Databases

2. Source Information:

(a) Creators:

-- 1. Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D.

-- 2. University Hospital, Zurich, Switzerland: William Steinbrunn, M.D.

-- 3. University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.

-- 4. V.A. Medical Center, Long Beach and Cleveland Clinic Foundation:

Robert Detrano, M.D., Ph.D.

(b) Donor: David W. Aha ([aha@ics.uci.edu](mailto:aha@ics.uci.edu)) (76) 856-8779

(c) Date: July, 1188

3. Past Usage:

1. Detrano,~R., Janosi,~A., Steinbrunn,~W., Pfisterer,~M., Schmid,~J.,

Sandhu,~S., Guppy,~K., Lee,~S., \& Froelicher,~V. (1189). {\it

International application of a new probability algorithm for the

diagnosis of coronary artery disease.} {\it American Journal of

Cardiology}, {\it 64},224--230.

-- International Probability Analysis

-- Address: Robert Detrano, M.D.

Cardiology 23-C

V.A. Medical Center

5901 E. 7th Street

Long Beach, CA 90020

-- Results in percent accuracy: (for 0.5 probability threshold)

Data Name: CDF CADENZA

-- Hungarian 77 74

Long beach 79 77

Swiss 81 81

-- Approximately a 77% correct classification accuracy with a

logistic-regression-derived discriminant function

2. David W. Aha & Dennis Kibler

--

-- Instance-based prediction of heart-disease presence with the

Cleveland database

-- NTgrowth: 77.0% accuracy

-- C4: 74.8% accuracy

3. John Gennari

-- Gennari, J.~H., Langley, P, \& Fisher, D. (1189). Models of

incremental concept formation. {\it Artificial Intelligence, 40},

3--61.

-- Results:

-- The CLASSIT conceptual clustering system achieved a 78.9% accuracy

on the Cleveland database.

4. Relevant Information:

This database contains 76 attributes, but all published experiments

refer to using a subset of 6 of them. In particular, the Cleveland

database is the only one that has been used by ML researchers to

this date. The "goal" field refers to the presence of heart disease

in the patient. It is integer valued from 0 (no presence) to 4.

Experiments with the Cleveland database have concentrated on simply

attempting to distinguish presence (values 1,2,3,4) from absence (value

0).

The names and social security numbers of the patients were recently

removed from the database, replaced with dummy values.

One file has been "processed", that one containing the Cleveland

database. All four unprocessed files also exist in this directory.

5. Number of Instances:

Database: # of instances:

Cleveland: 223

Hungarian: 214

Switzerland: 43

Long Beach VA: 120

6. Number of Attributes: 76 (including the predicted attribute)

7. Attribute Information:

-- Only 6 used

-- 1. #3 (age)

-- 2. #4 (sex)

-- 3. #9 (cp)

-- 4. #10 (trestbps)

-- 5. #4 (chol)

-- 6. #8 (fbs)

-- 7. #11 (restecg)

-- 8. #24 (thalach)

-- 9. #38 (exang)

-- 10. #40 (oldpeak)

-- 3. #41 (slope)

-- 4. #44 (ca)

-- 5. #51 (thal)

-- 6. #58 (num) (the predicted attribute)

-- Complete attribute documentation:

1 id: patient identification number

2 ccf: social security number (I replaced this with a dummy value of 0)

3 age: age in years

4 sex: sex (1 = male; 0 = female)

5 painloc: chest pain location (1 = substernal; 0 = otherwise)

6 painexer (1 = provoked by exertion; 0 = otherwise)

7 relrest (1 = relieved after rest; 0 = otherwise)

8 pncaden (sum of 5, 6, and 7)

9 cp: chest pain type

-- Value 1: typical angina

-- Value 2: atypical angina

-- Value 3: non-anginal pain

-- Value 4: asymptomatic

10 trestbps: resting blood pressure (in mm Hg on admission to the

hospital)

3 htn

4 chol: serum cholestoral in mg/dl

5 smoke: I believe this is 1 = yes; 0 = no (is or is not a smoker)

6 cigs (cigarettes per day)

7 years (number of years as a smoker)

8 fbs: (fasting blood sugar > 40 mg/dl) (1 = true; 0 = false)

9 dm (1 = history of diabetes; 0 = no such history)

10 famhist: family history of coronary artery disease (1 = yes; 0 = no)

11 restecg: resting electrocardiographic results

-- Value 0: normal

-- Value 1: having ST-T wave abnormality (T wave inversions and/or ST

elevation or depression of > 0.05 mV)

-- Value 2: showing probable or definite left ventricular hypertrophy

by Estes' criteria

12 ekgmo (month of exercise ECG reading)

13 ekgday(day of exercise ECG reading)

14 ekgyr (year of exercise ECG reading)

15 dig (digitalis used furing exercise ECG: 1 = yes; 0 = no)

16 prop (Beta blocker used during exercise ECG: 1 = yes; 0 = no)

17 nitr (nitrates used during exercise ECG: 1 = yes; 0 = no)

18 pro (calcium channel blocker used during exercise ECG: 1 = yes; 0 = no)

19 diuretic (diuretic used used during exercise ECG: 1 = yes; 0 = no)

20 proto: exercise protocol

1 = Bruce

2 = Kottus

3 = McHenry

4 = fast Balke

5 = Balke

6 = Noughton

7 = bike 70 kpa min/min (Not sure if "kpa min/min" is what was

written!)

8 = bike 45 kpa min/min

9 = bike 100 kpa min/min

10 = bike 75 kpa min/min

3 = bike 50 kpa min/min

4 = arm ergometer

21 thaldur: duration of exercise test in minutes

22 thaltime: time when ST measure depression was noted

23 met: mets achieved

24 thalach: maximum heart rate achieved

25 thalrest: resting heart rate

26 tpeakbps: peak exercise blood pressure (first of 2 parts)

27 tpeakbpd: peak exercise blood pressure (second of 2 parts)

36 dummy

37 trestbpd: resting blood pressure

38 exang: exercise induced angina (1 = yes; 0 = no)

39 xhypo: (1 = yes; 0 = no)

40 oldpeak = ST depression induced by exercise relative to rest

41 slope: the slope of the peak exercise ST segment

-- Value 1: upsloping

-- Value 2: flat

-- Value 3: downsloping

42 rldv5: height at rest

43 rldv5e: height at peak exercise

44 ca: number of major vessels (0-3) colored by flourosopy

45 restckm: irrelevant

46 exerckm: irrelevant

47 restef: rest raidonuclid (sp?) ejection fraction

48 restwm: rest wall (sp?) motion abnormality

0 = none

1 = mild or moderate

2 = moderate or severe

3 = akinesis or dyskmem (sp?)

49 exeref: exercise radinalid (sp?) ejection fraction

50 exerwm: exercise wall (sp?) motion

51 thal: 3 = normal; 6 = fixed defect; 7 = reversable defect

52 thalsev: not used

53 thalpul: not used

54 earlobe: not used

55 cmo: month of cardiac cath (sp?) (perhaps "call")

56 cday: day of cardiac cath (sp?)

57 cyr: year of cardiac cath (sp?)

58 num: diagnosis of heart disease (angiographic disease status)

-- Value 0: < 50% diameter narrowing

-- Value 1: > 50% diameter narrowing

(in any major vessel: attributes 59 through 68 are vessels)

59 lmt

60 ladprox

61 laddist

62 diag

63 cxmain

64 ramus

65 om1

66 om2

67 rcaprox

68 rcadist

69 lvx1: not used

70 lvx2: not used

71 lvx3: not used

72 lvx4: not used

73 lvf: not used

74 cathef: not used

75 junk: not used

76 name: last name of patient

(I replaced this with the dummy string "name")

9. Missing Attribute Values: Several. Distinguished with value -9.0.

10. Class Distribution:

Database: 0 1 2 3 4 Total

Cleveland: 84 55 36 27 5 223

Hungarian: 108 37 18 20 7 214

Switzerland: 8 48 24 22 5 43

Long Beach VA: 51 56 41 42 10 120

### Our Context

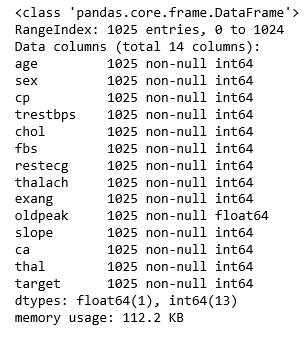
This data set consists of four databases: Cleveland, Hungary, Switzerland, and Long Beach V. It contains 76 attributes, including the predicted attribute, but all published experiments refer to using a subset of 14 of them. The "target" field refers to the presence of heart disease in the patient. It is integer valued 0 = no disease and 1 = disease.

### Our Content

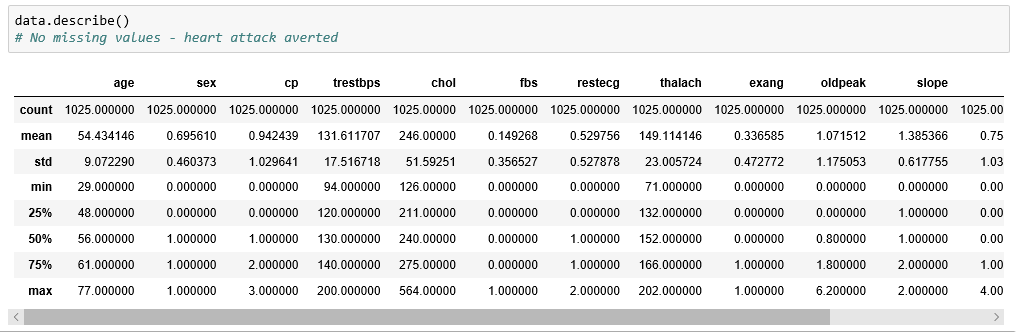
Attribute Information: > 1. age > 2. sex > 3. chest pain type (4 values) > 4. resting blood pressure > 5. serum cholestoral in mg/dl > 6. fasting blood sugar > 40 mg/dl > 7. resting electrocardiographic results (values 0,1,2) > 8. maximum heart rate achieved > 9. exercise induced angina > 10. oldpeak = ST depression induced by exercise relative to rest > 11. the slope of the peak exercise ST segment > 12. number of major vessels (0-3) colored by flourosopy > 13. thal: 0 = normal; 1 = fixed defect; 2 = reversable defect The names and social security numbers of the patients were recently removed from the database, replaced with dummy values.

**Methodology**

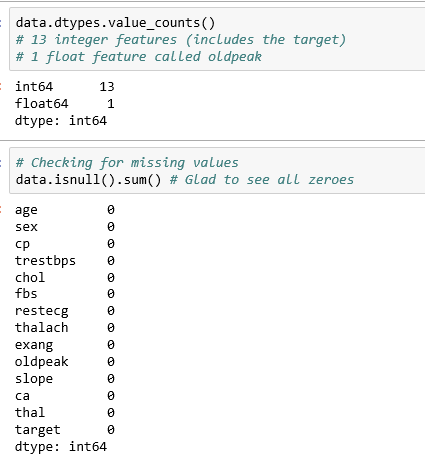
Exploratory Data Analysis



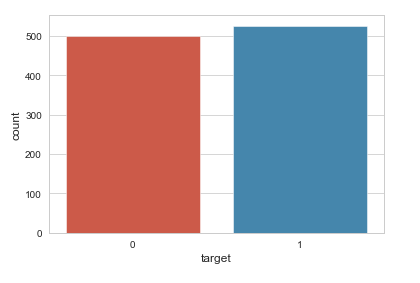
Dataset’s Basic Information



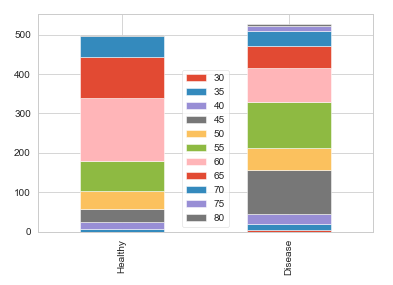
Dataset’s Description



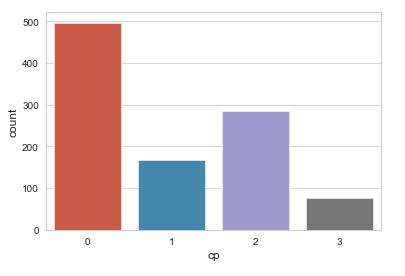
No missing values - heart attack averted



Target Varisable Distribution-Almost Balanced



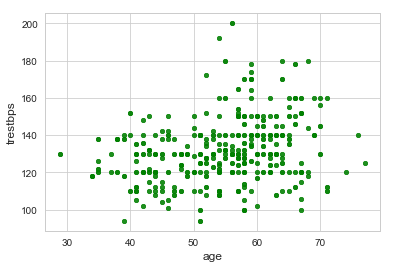
Age with respect to heart disease



Chest Pain type

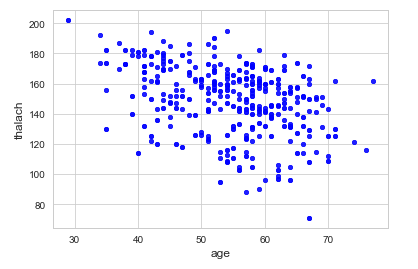
Value 1: typical angina -- Value 2: atypical angina -- Value 3: non-anginal pain -- Value 4: asymptomatic

In dataset, value starts from 0



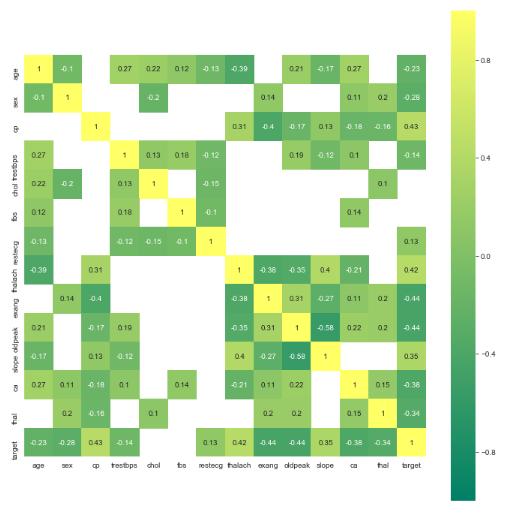
Relationship between age and trestbps

More people will have higher blood pressure as they age



Relationship between age and maximum heartrate achieved

As you age, the maximum heart rate you can achieve will gradually reduce



Correlation Matrix

oldpeak -0.438441

exang -0.438021

ca -0.381205

thal -0.257838

sex -0.199501

age -0.149236

trestbps -0.58772

restecg 0.54468

slope 0.26554

thalach 0.414895

cp 0.426854

**Data Modelling**

In reference to the research article Using recurrent neural network models for early detection of heart failure onset(2) as RNN gives the highest accuracy we used LSTM units to implement RNN in our model.

**Long short-term memory** (**LSTM**)(1) is an artificial [recurrent neural network](https://en.wikipedia.org/wiki/Recurrent_neural_network) (RNN) architecture used in the field of [deep learning](https://en.wikipedia.org/wiki/Deep_learning). Unlike standard [feedforward neural networks](https://en.wikipedia.org/wiki/Feedforward_neural_network), LSTM has feedback connections that make it a "general purpose computer" (that is, it can compute anything that a [Turing machine](https://en.wikipedia.org/wiki/Turing_machine) can). It can not only process single data points (such as images), but also entire sequences of data (such as speech or video). For example, LSTM is applicable to tasks such as unsegmented, connected [handwriting recognition](https://en.wikipedia.org/wiki/Handwriting_recognition) or [speech recognition](https://en.wikipedia.org/wiki/Speech_recognition). [Bloomberg Business Week](https://en.wikipedia.org/wiki/Bloomberg_Business_Week) wrote: "These powers make LSTM arguably the most commercial AI achievement, used for everything from predicting diseases to composing music."

A common LSTM unit is composed of a **cell**, an **input gate**, an **output gate** and a **forget gate**. The cell remembers values over arbitrary time intervals and the three *gates* regulate the flow of information into and out of the cell.

LSTM networks are well-suited to [classifying](https://en.wikipedia.org/wiki/Classification_in_machine_learning), [processing](https://en.wikipedia.org/wiki/Computer_data_processing) and [making predictions](https://en.wikipedia.org/wiki/Predict) based on [time series](https://en.wikipedia.org/wiki/Time_series) data, since there can be lags of unknown duration between important events in a time series. LSTMs were developed to deal with the exploding and [vanishing](https://en.wikipedia.org/wiki/Vanishing_gradient_problem) gradient problems that can be encountered when training traditional RNNs. Relative insensitivity to gap length is an advantage of LSTM over RNNs, [hidden Markov models](https://en.wikipedia.org/wiki/Hidden_Markov_models) and other sequence learning methods in numerous applications

There are several architectures of LSTM units. A common architecture is composed of a **cell** (the memory part of the LSTM unit) and three "regulators", usually called gates, of the flow of information inside the LSTM unit: an **input gate**, an **output gate** and a **forget gate**. Some variations of the LSTM unit do not have one or more of these gates or maybe have other gates. For example, [gated recurrent units](https://en.wikipedia.org/wiki/Gated_recurrent_unit) (GRUs) do not have an output gate.

Intuitively, the *cell* is responsible for keeping track of the dependencies between the elements in the input sequence. The *input gate* controls the extent to which a new value flows into the cell, the *forget gate* controls the extent to which a value remains in the cell and the *output gate* controls the extent to which the value in the cell is used to compute the output activation of the LSTM unit. The activation function of the LSTM *gates* is often the [logistic function](https://en.wikipedia.org/wiki/Logistic_function).

There are connections into and out of the LSTM *gates*, a few of which are recurrent. The weights of these connections, which need to be learned during [training](https://en.wikipedia.org/wiki/Supervised_learning), determine how the gates operate.

There have been several successful stories of training, in a non-supervised fashion, RNNs with LSTM units.

[Bill Gates](https://en.wikipedia.org/wiki/Bill_Gates) called it a “huge milestone in advancing artificial intelligence” when bots developed by [OpenAI](https://en.wikipedia.org/wiki/OpenAI) were able to beat humans in the game of Dota 2. OpenAI Five consists of five independent but coordinated neural networks. Each network is trained by a policy gradient method without supervising teacher and contains a single-layer, 1016-unit Long-Short-Term-Memory that sees the current game state and emits actions through several possible action heads.

[OpenAI](https://en.wikipedia.org/wiki/OpenAI) also trained a similar LSTM by policy gradients to control a human-like robot hand that manipulates physical objects with unprecedented dexterity.

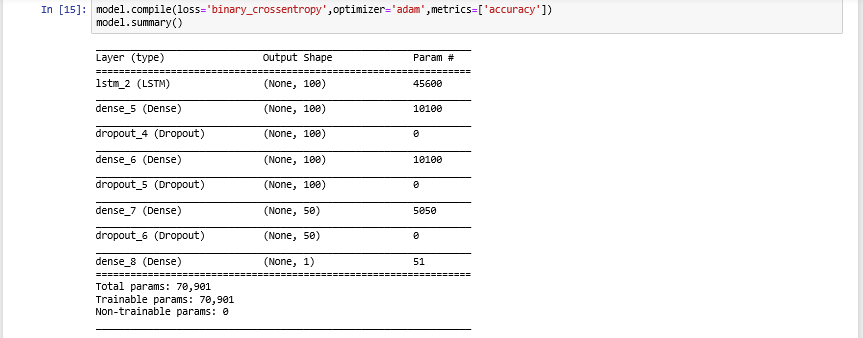
[DeepMind](https://en.wikipedia.org/wiki/DeepMind)'s program AlphaStar used a deep LSTM core to excel at the complex video game [Starcraft](https://en.wikipedia.org/wiki/Starcraft). This was viewed as significant progress towards Artificial General Intelligence.

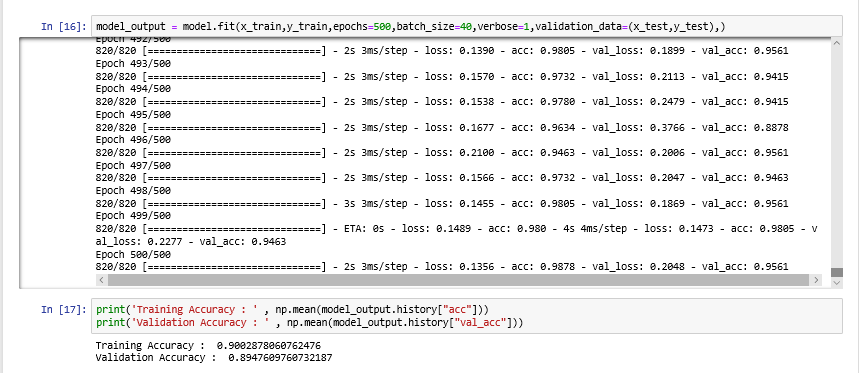
**Implementation**

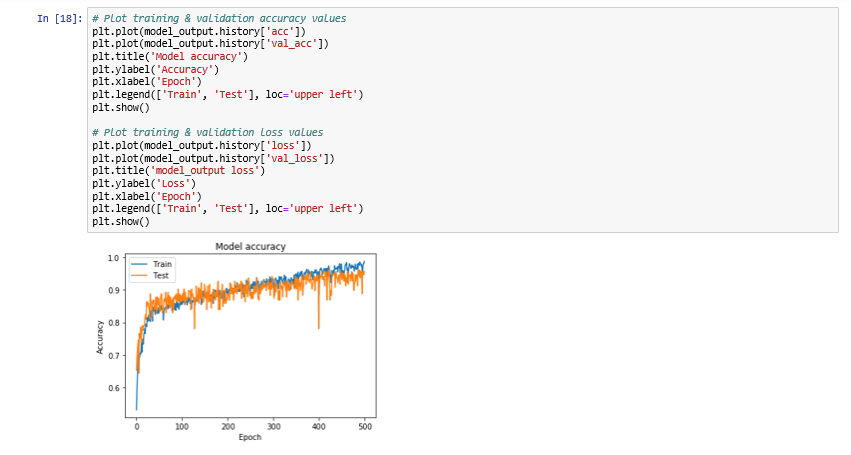
ModelCreation-Final.ipynb

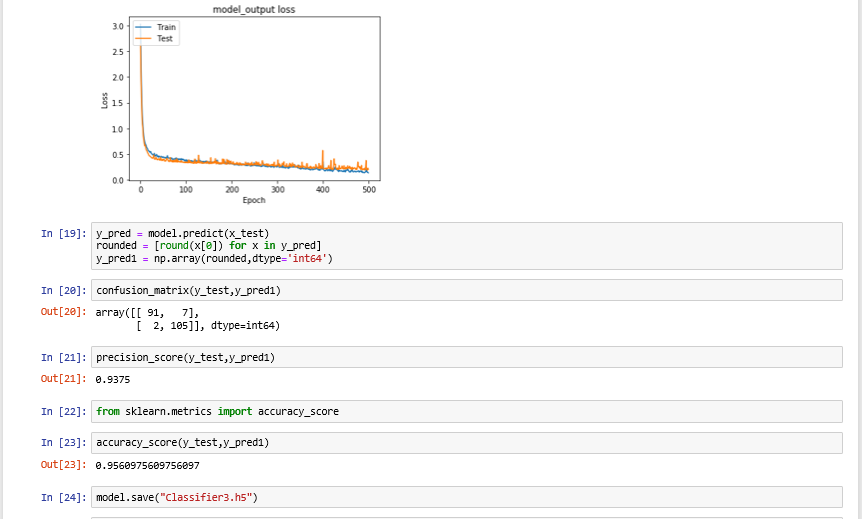




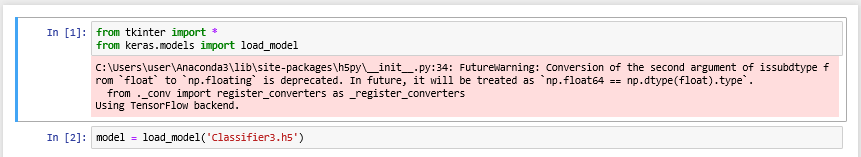


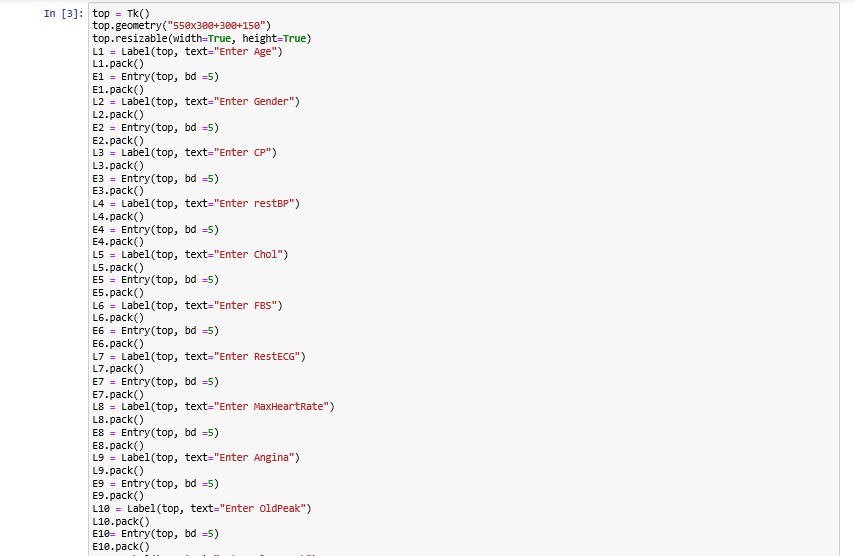


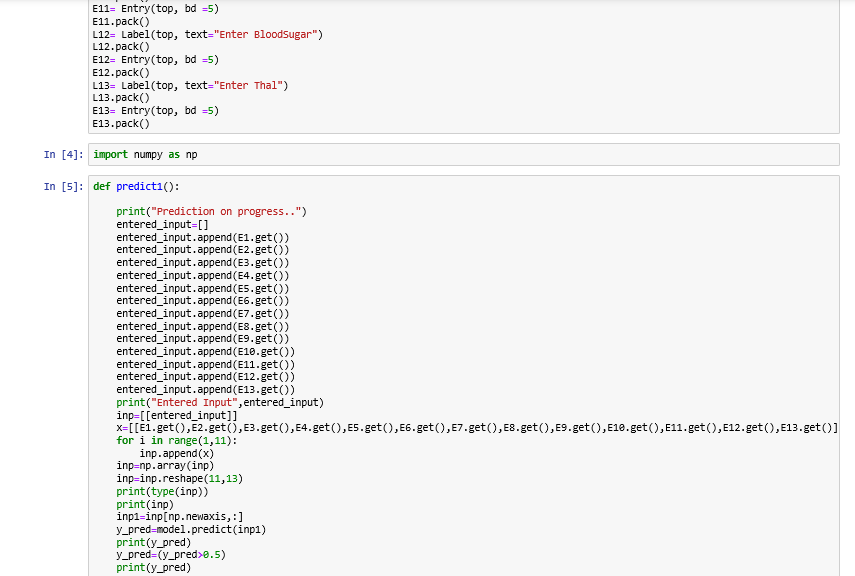


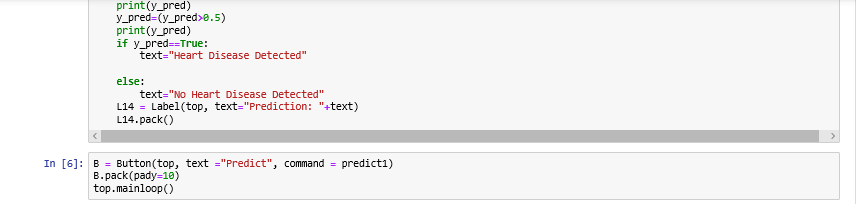


UI Generation-Model\_UI.ipynb

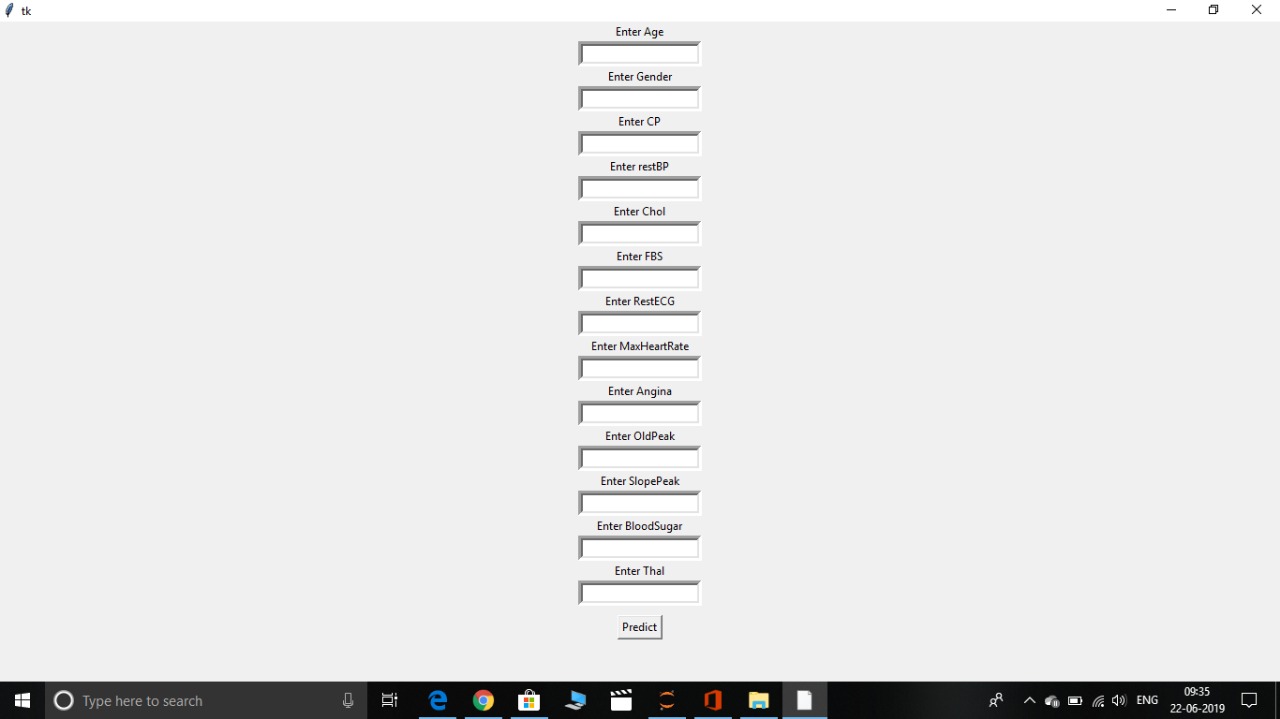


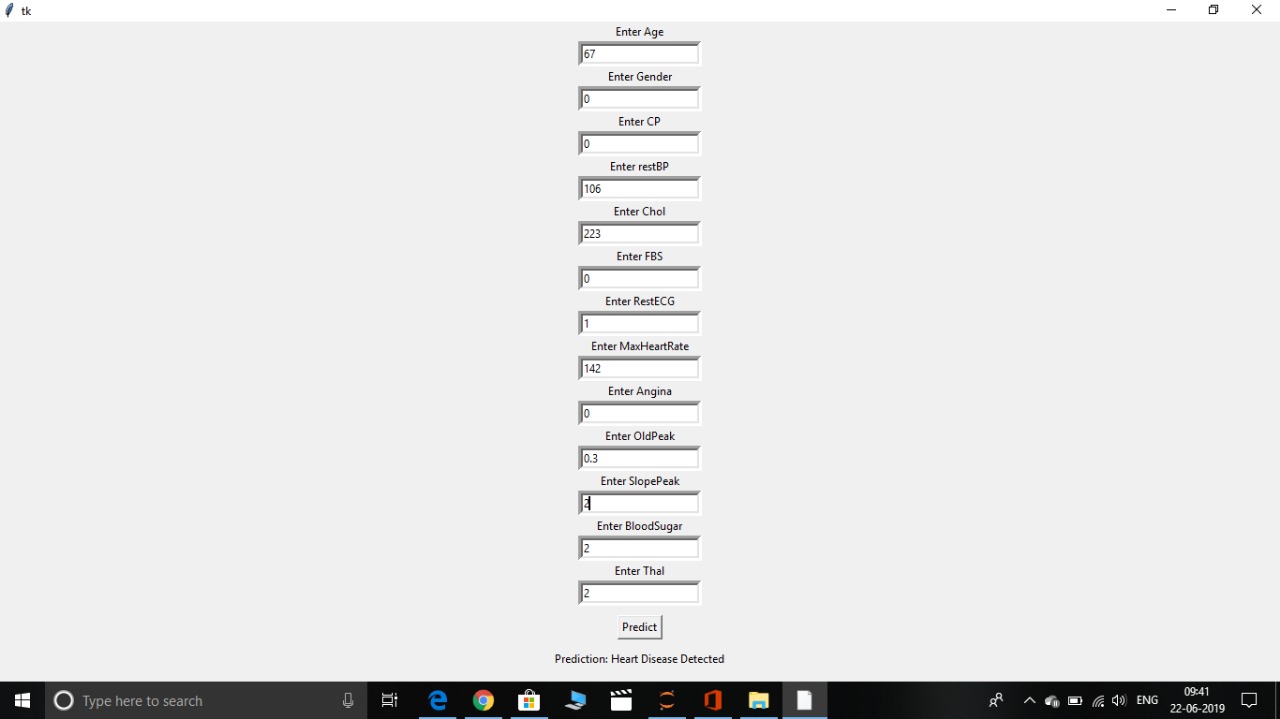
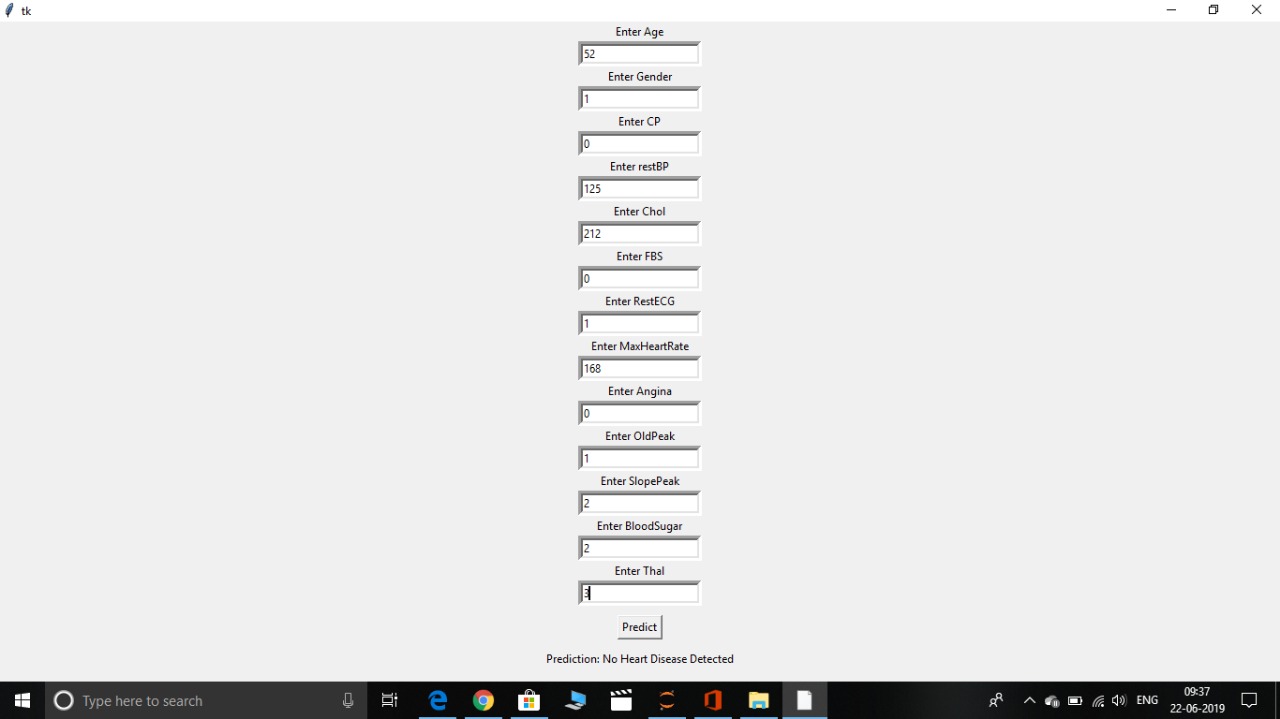






Output





**Suggestions**

-The User Interface Generated has to be improved to be more user friendly

-Other forms of RNN can be used for predictive modeling

**References**

1.<https://en.wikipedia.org/wiki/Long_short-term_memory>

2.[https://www.researchgate.net/publication/226096917\_Using\_recurrent\_neural\_network\_models\_for\_early\_detection\_of\_heart\_failure\_onset](https://www.researchgate.net/publication/306096925_Using_recurrent_neural_network_models_for_early_detection_of_heart_failure_onset)

3. Cho K, Van Merrienboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Empirical Methods in Natural Language Processing (EMNLP). 118:916–926. Doha, Qatar.

4. Hinton G, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. Neural Comput 1206;10(7):719–754.

5. Bengio Y. Learning deep architectures for AI. Foundations Trends Machine Learning. 1209;2(1):1–47.

6. Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems (NIPS) 116:226–36. Lake Tahoe, Nevada, United States.

7. Vincent P, Larochelle H, Bengio Y, Manzagol P-A. Extracting and composing robust features with denoising autoencoders. In International Conference on Machine learning (ICML) 1208:1096–223. Helsinki, Finland.

8. Le Q, Ranzato M, Monga R, et al. Building high-level features using large scale unsupervised learning. In International Conference on Machine Learning (ICML) 116, Edinburgh, Scotland, UK.

9. Lee H, Pham P, Largman Y, Ng A. Unsupervised feature learning for audio classification using convolutional deep belief networks. In Advances in Neural Information Processing Systems (NIPS) 1209;1096–224. Vancouver, British Columbia, Canada.

10. Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. Signal Process Mag 116;21(6):82–97.

11. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In arXiv preprint arXiv:501.3781 117.

12. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In Advances in Neural Information Processing Systems (NIPS) 117:243–259. Lake Tahoe, Nevada, United States.

13. Socher R, Pennington J, Huang E, Ng A, Manning C. Semi-supervised recursive autoencoders for predicting sentiment distributions. In Empirical Methods in Natural Language Processing (EMNLP). 115:71–81. Edinburgh, UK.

14. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1197;9(8):927–980.

15. Grosicki E, El Abed H. ICDAR 1209 handwriting recognition competition. In International Conference on Document Analysis and Recognition 1209:598–602. Barcelona, Spain.

16. Sak H, Senior A, Beaufays F. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In International Speech Communication Association 118;258–262. Singapore.

17. Zaremba W, Sutskever I, Vinyals O. Recurrent neural network regularization. In arXiv preprint arXiv:609.1521 118.

18. Luong M-T, Sutskever I, Le Q, Vinyals O, Zaremba W. Addressing the rare word problem in neural machine translation. In Association for Computational Linguistics (ACL) 119:3–11. Beijing, China.

19. Jozefowicz R, Zaremba W, Sutskever I. An empirical exploration of recurrent network architectures. In International Conference on Machine Learning (ICML). 119:1542–1550. Lille, France.

20. Lasko T, Denny J, Levy M. Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data. PloS One 117;8(6):e66261.

21. Che Z, Kale D, Li W, Bahadori M, Liu Y. Deep computational phenotyping. In Knowledge Discovery and Data Mining (KDD). 119:507–58. Sydney, NSW, Australia.

22. Hammerla N, Fisher J, Andras P, Rochester L, Walker R, Plotz T. PD disease state assessment in naturalistic environments using deep learning. In AAAI 119. 942–948. Austin, Texas, USA.

23. Lipton Z, Kale D, Elkan C, Wetzell R. Learning to diagnose with LSTM recurrent neural networks. In arXiv preprint arXiv: 73.03677 120.

24. Minarro-Gimenez J, Marin-Alonso O, Samwald M. Exploring the application of deep learning techniques on medical text corpora. Stud Health Technol Inform 117;117:584–588.

25. De Vine L, Zuccon G, Koopman B, Sitbon L, Bruza P. Medical semantic similarity with a neural language model. In International Conference on Information and Knowledge Management (CIKM). 118;1011–1014. Shanghai, China.

26. Choi Y, Chiu C, Sontag D. Learning low-dimensional representations of medical concepts. In American Medical Informatics Association on Clinical Research Informatics 120. San Francisco, CA.

27. Choi E, Schuetz A, Stewart W, Sun J. Medical concept representation learning from electronic health records and its application on heart failure prediction. In arXiv preprint arXiv:802.03686 120