Hybrid Learning Approaches

Disease Prediction and Stock Prediction

Shyamnath Premnadh
Barshana Banerjee
Debanjali Biswas
Mst. Mahfuja Akter

Supervisor: Maria Maleshkova

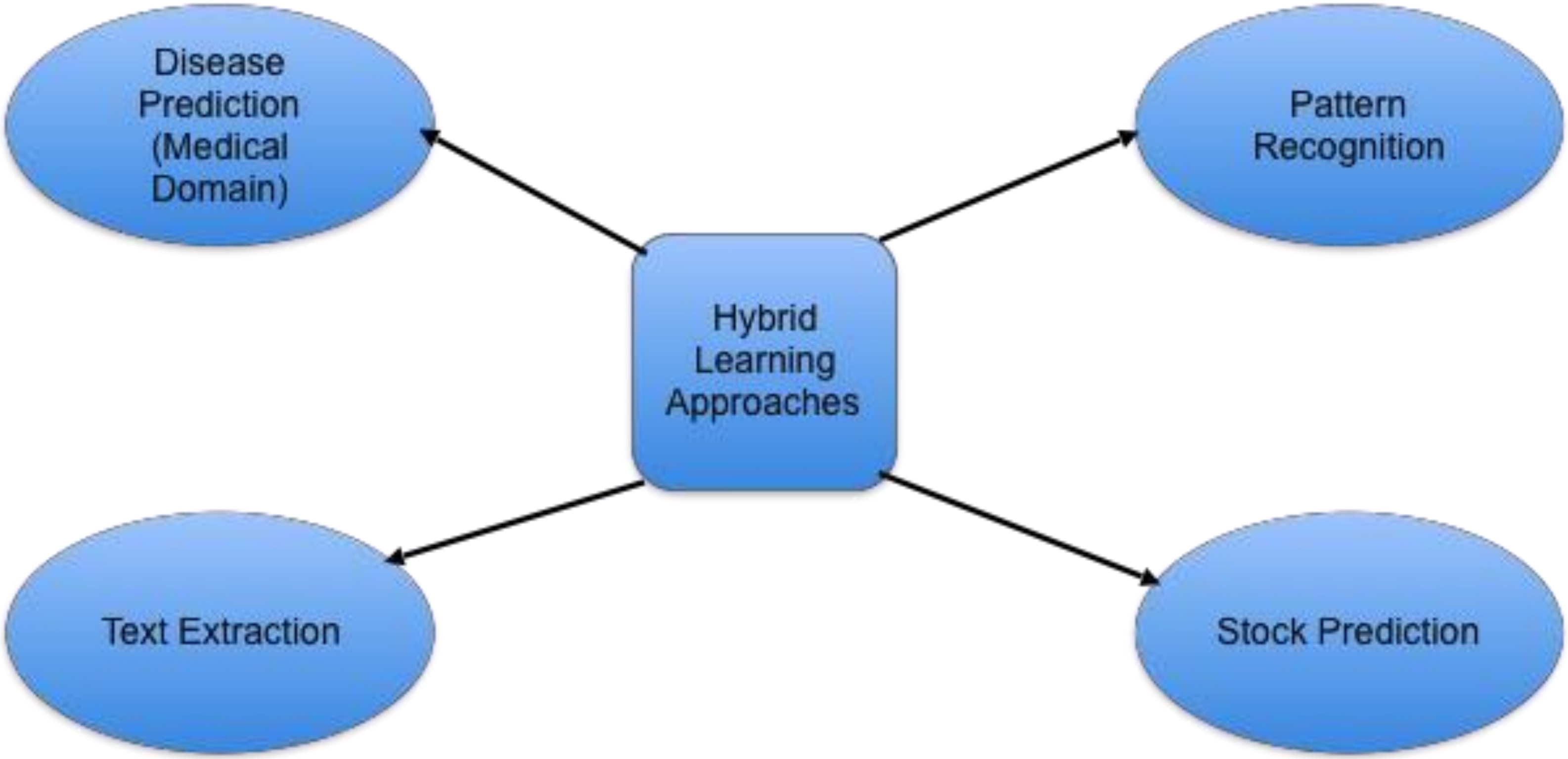
Introduction

Hybrid learning algorithm is a combination of learning paradigms that provide advantages that the original paradigms do not possess.

Advantages include:

- Better efficiency over the single calibrated learning algorithm
- Solve the complex problem that is difficult to be solved by individual learning algorithm
- Compensate the limitation of each single algorithm
- Share the complementary benefits from each combined algorithm

Use Case / Motivation Scenario/Problem



Recap

- Pattern Recognition:
  - The Hybrid method for Pattern Recognition provides an automated means to train the BP network using the KFLANN output as the teacher value.
  - Benefits:
    - Cost effective way learning model
    - KFLANN shares the interpolation capabilities with BP
    - Classification performance is not compromised
- Text Extraction:
  - The novelty of this model was the two techniques have never been combined for an IE system before.
  - The results showed that the use of HMM compensated the relatively low performance of less adequate classifiers and feature sets.
  - Makes it possible to extend to other models.

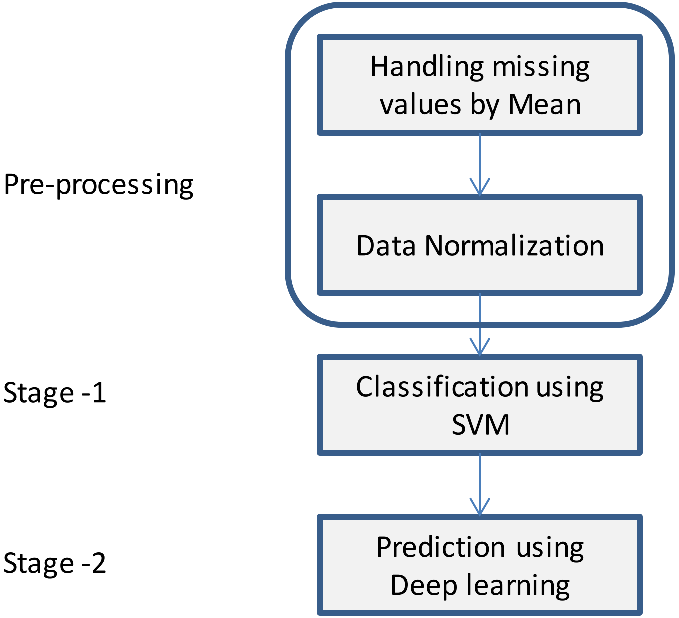
Disease prediction

Referred from : Deeksha Kaul, Harika Raju and B.K. Tripathy

Comparative Analysis of Pure and Hybrid Machine Learning Algorithms for Risk Prediction of Diabetes Mellitus

Hybrid Deep Learning Method

- Support Vector machine
- Deep Neural Network
- Hybrid Deep Learning

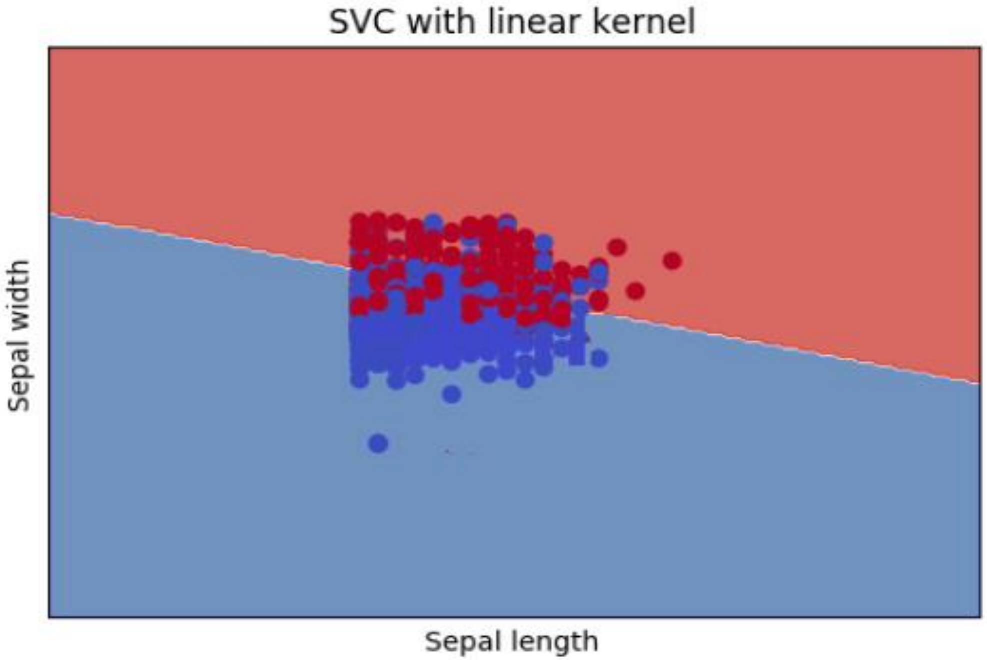


Dataset (Pima Indians Diabetes dataset)

Sl No	Name of Attribute
1	No. of Pregnancies
2	Plasma Glucose Concentration
3	Diastolic Blood Pressure (mm Hg)
4	Triceps Skin Fold Thickness (mm)
5	2-hour Serum Insulin (mu U/ml)
6	Body Mass Index
7	Diabetes Pedigree Function
8	Age
9	Class Variable (0 or 1)

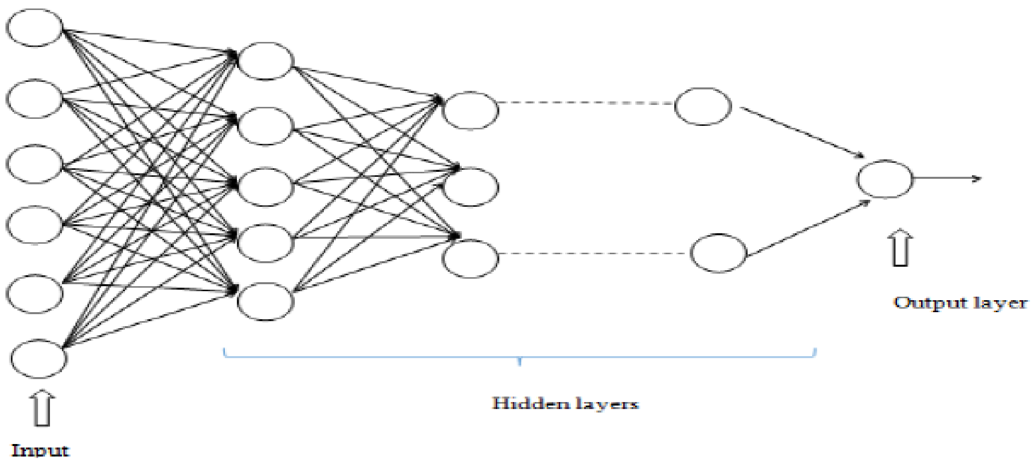
- To avoid over fitting, the data set has been divided by training set, cross-validation set and test set in the ratio of 6:2:2
- By using ‘plasma glucose concentration’, they predict the progression of diseases according to the guidelines set by American Diabetes Association.

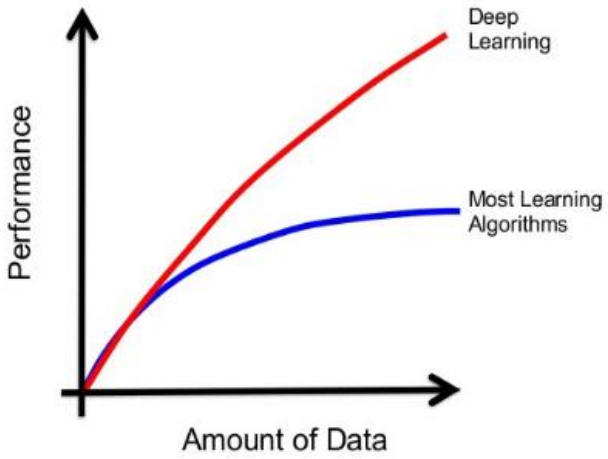
Support Vector Machine(SVM)



- A supervised machine learning approach.
- Separates data into two clusters (positive and negative) by a hyperplane.
- Using kernel trick, it can separates linearly inseparable data also.

Deep Neural Network





- An unsupervised learning approach inspired by linking and functionalities of neurons in human brain.
- Trained the system with stochastic gradient descent using back-propagation.
- Gives better performance than most learning algorithms even with large data.

Confusion Matrix

	True positive	False positive	True negative	False negative
SVM	140	46	454	128
Deep learning	150	50	450	118
Hybrid deep learning	187	70	430	81

True positive 24.35%	False Negative 10.55%
False positive 09.11%	True Negative 55.99%

Model performance

Accuracy: (TP+TN)/(TP+FN+FP+TN)

RMSE: sqrt((FP+FN)/ TP+FN+FP+TN)

Precision: TP/(FP+TP)

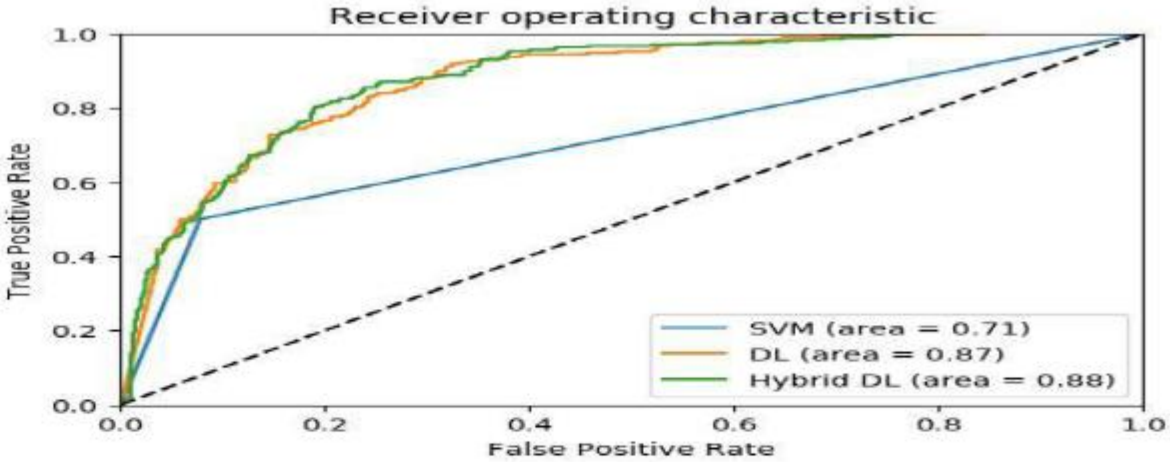
Performance analysis & Disease Progression

	SVM	Deep Neural Network	Hybrid Deep learning Network
Accuracy	0.7734	0.7812	0.8034
AUC	0.7152	0.8538	0.8733
MSE	0.2266	0.1489	0.1376
Precision	0.7526	0.7502	0.7625

Plasma glucose concentration	Stage
4.5 – 5	Stage 1
5.1 – 6.5	Stage 2
6.6 – 16	Stage 3
16.1 – 22	Stage 4
> 22	Stage 5

Claim

- Hybrid deep learning method provides the most satisfactory prediction of disease with least error rate, highest accuracy, precision and area under ROC curve compared to SVM and Deep learning network.



Future work

- The proposed algorithm is tested with only female data. Method should explore male data also.
- Han, Jianchao, Juan C. Rodriguez, and Mohsen Beheshti.: Diabetes data analysis and prediction model discovery using rapidminer. In: Second International Conference on Future Generation Communication and Networking, FGCN'08. vol. 3. IEEE(2008) --- This papers also worked on same dataset, by using Decision tree and ID3 algorithm which have 72% and 80% accuracy respectively.

Stock prediction

Referred from : Honghai Yu , Haifei Liu

Improved Stock Market Prediction by Combining Support Vector Machine and Empirical Mode Decomposition

Stock Prediction

- Predicting the direction of stock accurately is very crucial for investors to maximize profit.
- But, predicting the stock market is one of the most difficult things to do.
- Is hybrid learning a potential game changer in this domain?

Proposed Hybrid Model

The proposed method is a ***hybrid learning algorithm*** which combines:

- **Empirical Mode Decomposition (EMD)** - to analyze non-stationary time series data.
- **Support Vector Machine (SVM)** – prediction by using different kernel functions and learning parameters of SVMs.

Running Example

- Step 1 : Decomposition of the series with the help of EMD into sum of IMFs(intrinsic mode functions)

- Step 2 : Regression estimation using SVMs

- Step 3 : Combine predicted results

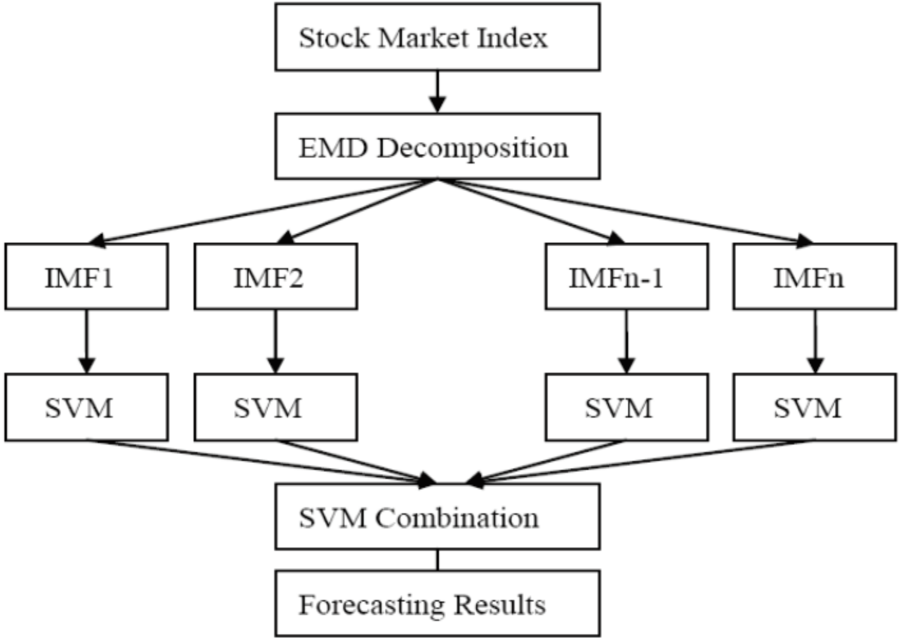


Figure 1. The Process of SVM-EMD model

Formal Definition of the Approach/ Methods

- Single SVM model is not appropriate for dynamic time series data.
- EMD helps to convert non stationary data to stationary form by decomposing it into different intrinsic mode functions (IMFs).
- SVMs are constructed for each IMF by finding appropriate kernel function to forecast the market index.

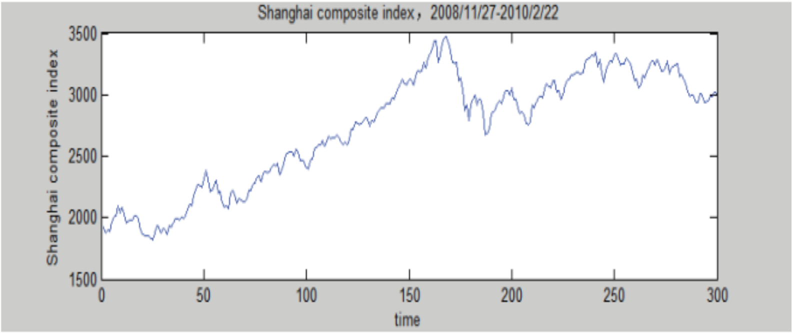


Figure 2. The trend of Shanghai Composite Index

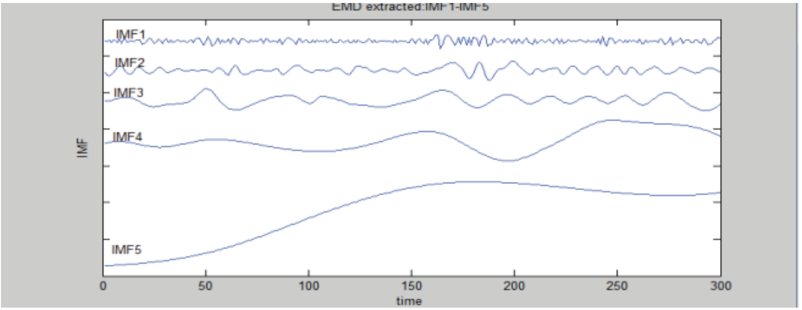


Figure 3. IMFs of Shanghai Composite Index through EMD

Evaluation

- Single SVM prediction:- Comparative error is from 0.11% to 10.87%

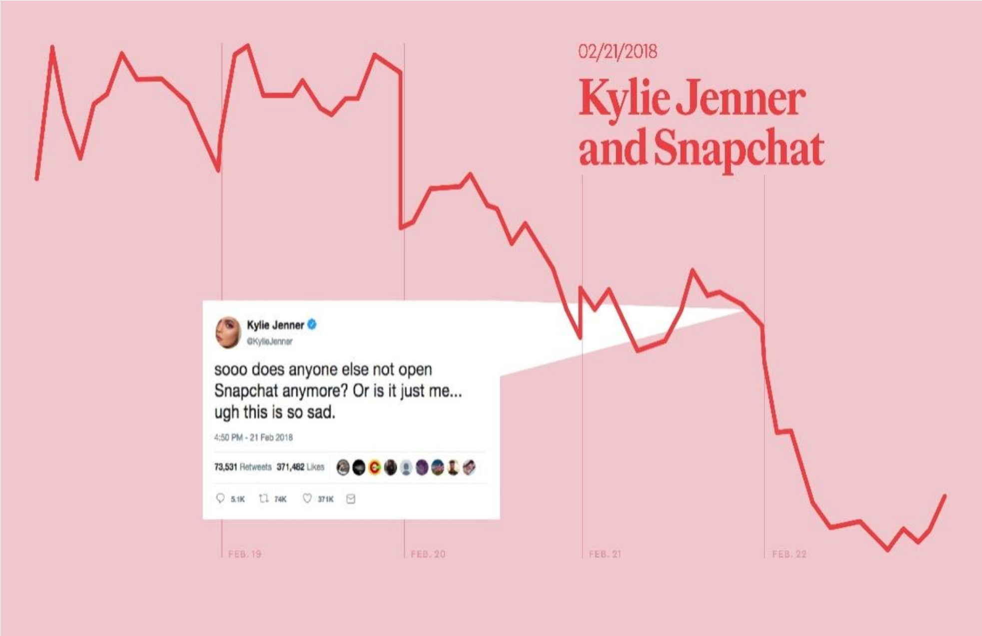
- EMD-SVM model :- Comparative error is from 0.10% to 1.91%.

	SVM Prediction Results			EMD-SVM Prediction Results		
Actual Data	Prediction	Absolute Error	Comparative Error	Prediction	Absolute Error	Comparative Error
2982.58	3030.20	47.62	1.60%	2957.15	25.43	0.85%
3022.18	3025.52	3.34	0.11%	2974.68	47.50	1.57%
3060.62	3004.67	55.95	1.83%	3010.18	50.44	1.65%
3051.94	2968.92	83.02	2.72%	3048.86	3.08	0.10%
3087.84	2922.53	165.31	5.35%	3076.79	11.05	0.36%
3073.11	2871.43	201.68	6.56%	3089.97	16.86	0.55%
3097.01	2821.76	275.25	8.89%	3090.45	6.57	0.21%
3023.37	2778.49	244.87	8.10%	3081.10	57.73	1.91%
3031.07	2744.68	286.39	9.45%	3064.34	33.27	1.10%
3053.23	2721.29	331.94	10.87%	3042.51	10.72	0.35%

Conclusion & Future work

- If various political, social & economic factors also affect stock market other than the technical indicators.

- Snapchat faced a loss of $1.3 billion stock after Kylie Jenner tweet.



Adapted from Tweets that Turned the Stock Market Upside Down
By Adam Kornblum, [image] Available at: https://www.ogilvy.com/feed/11-tweets-that-turned-the-stock-market-upside-down/

Thank You!