Recurrent Neural Networks with Non-Sequential Data to Predict Hospital Readmission of Diabetic Patients

Chahes Chopra
Department of Information Technology
ABV-Indian Institute of Information Technology and Management, Gwalior
+91-7354732013
ipg_2014031@iiitm.ac.in

Shivam Sinha
Department of Information Technology
ABV-Indian Institute of Information Technology and Management, Gwalior
+91-7770806859
ipg_2014082@iiitm.ac.in

Shubham Jaroli
Department of Information Technology
ABV-Indian Institute of Information Technology and Management, Gwalior
+91-8889857636
ipg_2014085@iiitm.ac.in

Anupam Shukla
Department of Information Technology
ABV-Indian Institute of Information Technology and Management, Gwalior
+91-957504800
dranupamshukla@gmail.com

Saumil Maheshwari
Department of Information Technology
ABV-Indian Institute of Information Technology and Management, Gwalior
+91-957504800
saumlimaheshwari@yahoo.co.in

ABSTRACT
Hospital readmissions are recognized as indicators of poor quality of care, such as inadequate discharge planning and care coordination. Moreover, most experts believe that many readmissions are unnecessary and avoidable. In the present paper, we design a Recurrent Neural Network model to predict whether a patient would be readmitted in the hospital and compared its accuracy with basic classifiers such as SVM, Random Forest and with Simple Neural Networks. RNN showed highest prediction power in all the models used and thus this can be used by hospitals to target high risk patients and prevent recurrent admissions.

CCS Concept
• Computing methodologies → Neural networks

Keywords
Recurrent neural networks; SVM; Random Forest; Hospital readmissions; Neural Networks.

1. INTRODUCTION
A hospital readmission [1] happens when a patient within a specified time interval, who had been discharged from a hospital is admitted again. Readmission can occur for any planned or unplanned reason. These factors pushed Hospital organizations to think about controlling readmission rates as increase in readmission rates leads to increase in penalties. There has been an increased usage of Readmission rates as an outcome measure in health services research and as a quality benchmark for health systems.

For Medicare patients, hospitalizations is very stressful and it is even more when they result in subsequent readmissions. In the United States, the Centres for Medicare and Medicaid Services (CMS) has reported [2] that 76% of readmissions that occur were potentially avoidable. A lot of research studies proved that hospitals can be engage in several activities like clarifying patient discharge instructions, coordinating with post-acute care providers, vibrant cleaning mechanism to reduce the rate of readmissions of patients. But these individual follow ups can be costly. Therefore this also raises a big question that which patient groups must be targeted to effectively use available resources for preventing readmission. Models that can predict accurately these are of a great help for hospitals all over the world as they can put extra efforts on high risk patients and can decrease their readmission rates.

The goal of the paper is to: 1) develop an accurate and generalized machine learning model that is applicable to predicting 30-day readmission for diabetic patients 2) To demonstrate the efficiency of Recurrent Neural Network in non-sequential data. 3) Compare RNN results with the results of other supervised algorithms such as SVM, Random Forest etc.

The paper proceeds as follows: Section 2 presents the literature review done about the basic models deployed prior to this study. Section 3 introduces the dataset we use, the data pre-processing and information about Recurrent Neural Networks. Section 4 reports the prediction results, compares the results achieved from different models and estimates prediction confidence.

2. LITERATURE REVIEW
During the last decade several research work is carried out in the field of healthcare. Hospital readmission prediction is important and crucial application of healthcare for the improved quality of life of an individual.

Author [3] examined facility-level rates of all-cause, unplanned hospital readmissions for 30 days following discharge from inpatient rehabilitation facilities. They used the dataset of

ICCBB 2017, October 18–20, 2017, Newark, NJ, USA
2017 Association for Computing Machinery.
ACM ISBN 978-1-4503-5322-9/17/10…$15.00
https://doi.org/10.1145/3155077.3155081

beneficiaries who were discharged from IRFs in 2013–2014. This study found that the national observed hospital readmission rate by 30 days post-IRF discharge was 13.1%. The mean IRF unadjusted readmission rate was 12.4% (SD = 3.5%) and the mean risk-standardized rate was 13.1% (SD = 0.8%). In [4] author's objective was to determine whether receipt of therapy and number and timing of therapy visits decreased hospital readmission risk in stroke survivors discharged home. They showed that during the first 30 days after discharge, 31% of patients saw a therapist in the home, 11% saw a therapist in an outpatient setting, and 59% did not see a therapist. Relative to patients who had no therapist contact, those who saw an outpatient therapist were less likely to be readmitted to the hospital (odds ratio 0.73 [059-0.90]). In [5] authors developed a hospital readmission predictive model, which enables controlling the trade-off between reasoning transparency and predictive accuracy, by taking into account the unique characteristics of the learned database. A boosted C5.0 tree, as the base classifier, was ensemble with a support vector machine (SVM), as a secondary classifier. This study showed that The SVM predictions are characterized by greater sensitivity values (true positive rates) than are the C5.0 predictions, for a wider range of cut off values of the ROC curve, depending on a predefined confidence threshold for the base C5.0 classifier.

3. MACHINE LEARNING MODELS

3.1 Logistic Regression

It is the most basic model used in machine learning to classify data and things. It is named logistic after the function which is used at its base, Logistic function. Logistic or Sigmoid function

[6] was used first to describe the properties of the population growth in ecology. Using MLE this algorithm is optimized to give best coefficient for our training dataset. Like in Linear Regression, in Logistic Regression also some assumptions are made to give robust performance like Binary output variable and no noisy training dataset by assuming no outliers and misclassified instances.

Using MLE this algorithm is optimized to give best coefficient for our training dataset. Like in Linear Regression, in Logistic Regression also some assumptions are made to give robust performance like Binary output variable and no noisy training dataset by assuming no outliers and misclassified instances.

3.2 Support Vector Machines (SVM)

SVM or Support Vector Machines [7] are a type of supervised learning models which can perform non-linear classification using Kernel trick. In this trick function takes low dimensional space input and transform it into higher dimensional space units. SVM [8] can be used for both classification and regression. SVMs are based on the concept of Hyperplane which divides entire dataset into two parts. We can use SVM in python using scikit library which contain SVM as an inbuilt function. But it has some drawbacks also when dataset is very large it does not perform well as required training time is high.

3.3 Decision Tree

For predictive type of modelling, Decision Tree is a very old algorithm. Random Forest is one of the powerful variations of this algorithm. Its modern day name is CART i.e. Classification and Regression Trees. CART model is represented in the form of binary tree. The root node of the Decision Tree represents a single input variable (X) and the leaf node represents an output variable which is used to make predictions. Input variables and split points on the variables play a major role in formation of Binary Tree. For selecting variables to be use and the specific split or cut-points optimally, Greedy algorithm is used to minimize cost function.

This is called "Recursive Binary Splitting".

3.4 Random Forest

It is a type of ensemble method used for classification and regression. It follows Decision Tree (CART) model. Random Forest [9] is a forest of many decision trees. In this method there are randomness at both Row and Column level. At Row level, each of the chosen decision trees gets a random sample data from whole dataset and training occur differently for each tree. At Column level also random selected columns are only passed from all columns. RF is a combination of "stochastic discrimination "with Bagging and random feature selection with controlled variance. It is very effective as fast training of training set occurs but it also has some drawbacks as it is a predictive model rather than a descriptive model and they also doesn't give a lot of insight, hence other approaches are preferred when run time performance is important over random forest.

3.5 Artificial Neural Network (ANN)

Inspired by real life Biological Neuron Networks modern day ANN's are developed. From ANNs one can easily model and process nonlinear relationship between inputs and outputs in parallel. In order to improve the efficiency this algorithm dynamically updates the 'weights' present between each neuron. An appropriate cost function must be chosen for good efficiency. Cost function must learn to predict optimal solution to the problem. Simple neutral network is made up of 3 layers. The input layer is the first layer which is followed by a hidden layer and the outer layer is the last one. There may be one or more than one neurons in each layer. By increasing the number of hidden layers or number of neurons or Number of paths between layers a model can be made complex [10].

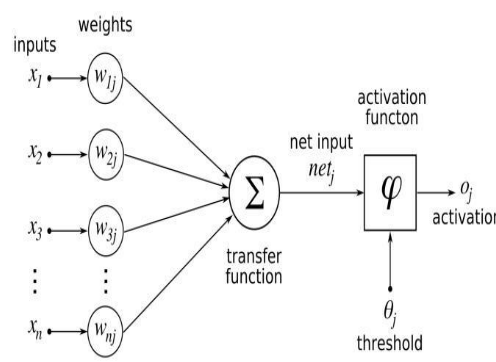

Figure 1. Artificial Feed Forward Network.

3.6 Recurrent Neural Networks (RNN)

To generate a new state vector in Recurrent Neural Network, input vector is combined with their state vector with the help of a fixed and learned function. In the language of the programming this can be stated as generating output from a fixed program which is given some input and some variable(internal). Recurrent Neural Networks are also called Turing complete as with proper weights they can be used as an arbitrary program simulator.

4. DATA AND PREPROCESSING

4.1 Dataset

We used the Health Facts database (Cerner Corporation, Kansas City, MO), a national data warehouse that collects comprehensive clinical records across hospitals throughout the United States. The Health Facts data we used was an extract representing 10 years (1999–2008) of clinical care at 130 hospitals of US hospitals and integrated delivery networks. The dataset was created in two steps.

First, important features were drawn out from the database. It was found that there were a total of 55 features which should be used for the study.

UCI Machine Learning Repository provides the above mentioned dataset.

Second, data was Pre-processed so to extract only the information that is useful for the research. Both steps are described in the following subsections.

Information was extracted from the database for encounters that satisfied the following criteria.

(1)It should be a case of hospital admission.

(2)Only diabetic patients are considered for the database.

(3)Stay duration of patient in the hospital lie in the range of 1-14 days.

(4)Laboratory tests were performed during the encounter.

(5)Medications were provided during the encounter.

101,766 encounters were identified to fulfil all of the above five inclusion criteria and were used in further analysis.

4.2 Data Preprocessing

4.2.1 Cleaning Of Data

Firstly in order to ensure that each patient has unique id we remove the record of those patients which appear more than one time. After that we remove the feature “patient_id, encounter_id” which tells about the Patient Number and Unique Identifier of a patient respectively.

4.2.2 Removing Biasness

To remove biasness we remove data of patients that are either dead or discharge to hospice. Also we duplicate some rows because number of patients admitted within 30 days are very low compare to those admitted after 30 days and the patients that are never admitted.

4.2.3 Feature Selection

There are total 55 features in this dataset, out of which 23 are medicine related features. After visualizing dependency of each feature with the readmission rate, we found that the medicines have the least role to play in the readmission, so 22 out of 23 medicine features were removed.

4.2.4 Missing Values

Some features, like “Weight” have 97% missing values and therefore cannot be used for analysis. We removed features having more than 30 percent of missing values. There is a feature, “medical specialty” that defines the specialty of attending physician which has some missing data so we fill “Missing” in the missing place as this is the important feature for analysis. After that we convert all the string categorical data into Integer categorical data to do our analysis.

4.2.5 Normalization of Dataset

We use L2 normalization [11] and Z score normalization in our dataset to remove the biasness among the features. Normalization brings the data on a common scale. It is used to standardize the range of independent variables or features of data, also known as feature scaling.

4.2.6 L2 Normalization

Let a matrix “A”:-

A = [a11, a12]
[a21, a22]

So for L2 normalization we calculate

a1 = sqrt (a11*a11+a21*a21)
a2 = sqrt (a12*a12+a22*a22)

And “A” becomes:-

A = [a11/a1, a12/a2]
[a21/a1, a22/a2]

1.1.1.1 Z-score Normalization

Let a matrix “A”:-

A = [a11, a12]
[a21, a22]

So in Z score normalization we have

Anorm = (A – mean(A))/std(A)*

*where std stands for Standard Deviation.

5.1 Recurrent Neural Network (RNN)

If we have some kind of sequential information or dataset and we want to apply Deep Learning methods on that dataset then first thing which comes to the mind is RNN [12]. In our evergreen traditional neural network inputs are independent of each other but in Recurrent Neural Network as the name “Recurrent” they perform the repetitive task for every sequence, with output being dependent on previous computations. This is done by internal state of RNN which contain context information also called its memory. This mechanism allow RNN to exploit a dynamically changing contextual window over the input sequence history.

But RNN has also showed great results in cases of non-sequential input information, for instance, Image captioning [13], where the image is a single non-sequential data point. It must be realized that powerful formalism of Recurrent Neural Network can be used in that situation also when input and outputs that were given are fixed vectors to process them in a sequential order. Even our data is not in the form of sequences, we can still develop powerful models and make them able to learn so that they process data in sequential order only.

In our study, we ran the model with 2 hidden layers and for 501 epochs. After 500 epochs, the validation error started to increase showing the chances of overfitting. Batch side was fixed to 13[14].

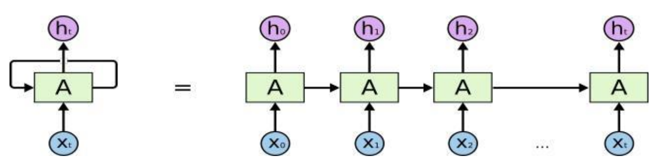

Figure 2. Recurrent Neural Network.

5 RESULTS

This section presents the readmission prediction results obtained from different predictive models. We will compare the prediction accuracy in terms of two factors AUC under ROC (Receiver Operating Characteristic) Curve, and overall accuracy.

Both area under ROC curve [15][16] and overall accuracy depends on the classifier's ability to rank patterns for positive class, but in the case of overall accuracy, it also depends on the ability to calculate the threshold in the ranking to classify the positive class.

To train and test the distinct models the data set is divided into a Train data set and a Test data set. The used proportion was 80% of data going to training and the remaining 20% to test.

Applying different algorithms on the dataset, with default parameters, we obtained results as mentioned in Table 1.

Table 1. Comparison among various models

Method	Area Under ROC Curve	Accuracy
Logistic Regression	0.53	62.91%
SVM	0.56	64.88%
Decision Tree	0.68	71.56%
Random Forest	0.73	74.64%
Simple Neural Network (2-layer)	0.61	69.53%
Recurrent Neural Network (2-Layer)	0.80	81.12%

It was found that Logistic Regression and SVM did a poor job in predicting the readmissions with accuracy of less than 65% each. Decision Tree and Random Forest performed better than the previous models but the accuracy was still unappreciable (71.56% and 74.64% respectively).

Simple Neural Network model couldn't perform well either and gave the accuracy of 69.53% with ROC of 0.61.

The best model was Recurrent Neural Network, which was able to predict the instances with an accuracy of up to 81%. The ROC for the same was 0.80.

K-fold cross validation is used to test and evaluate the algorithms. In this process the data is divided into K subsets. As the test set one of the subsets is used and from the training set the other subsets, i.e., K-1 is used each time. All 'K' trials are calculated from performance statistics. How well a classifier can perform on unseen data can be shown easily through this method.

To ensure that RNN model did not over fit, we divided the dataset into 3 parts: 75% for training and 15% for validation and 10% for testing. Then we calculated the validation loss and training loss of the respective dataset parts. Results showed that validation error was low, but slightly higher than the training error, thus showing that the model did not over fit.

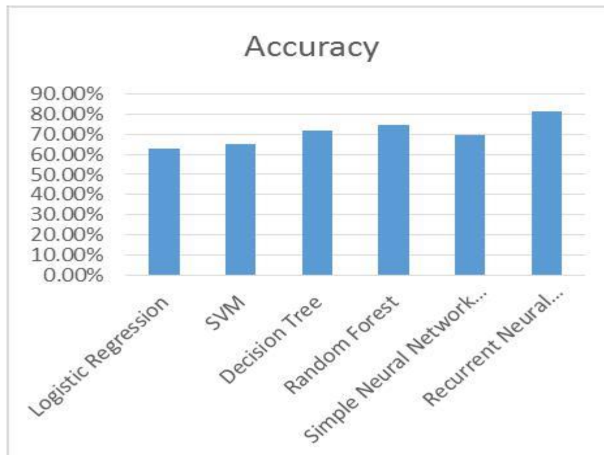

Figure 3. Methods vs accuracy.

Curves showing ROC of different models

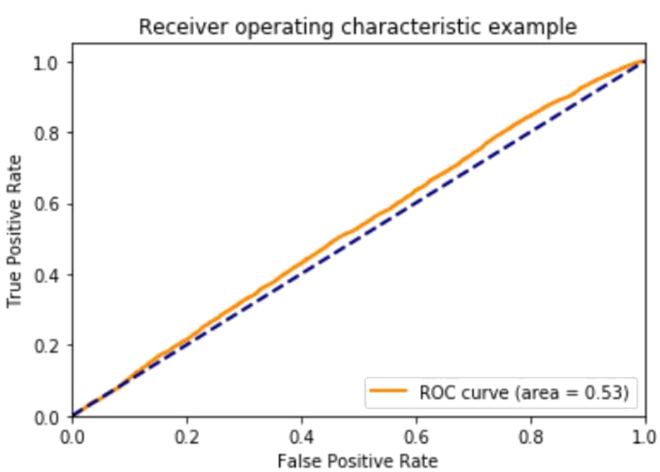

Figure 4. Logistic regression.

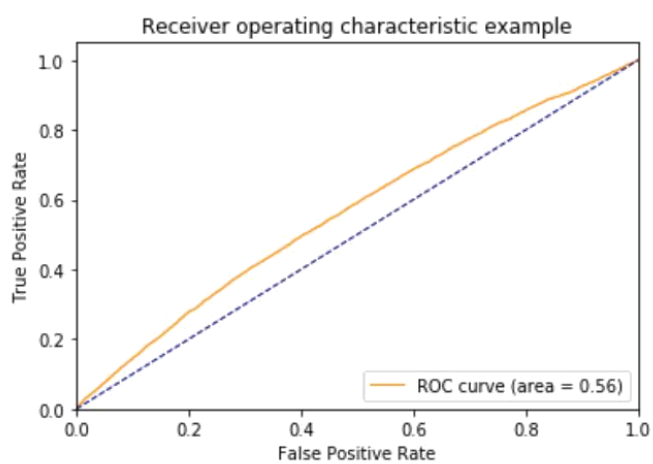

Figure 5. Support vector machines.

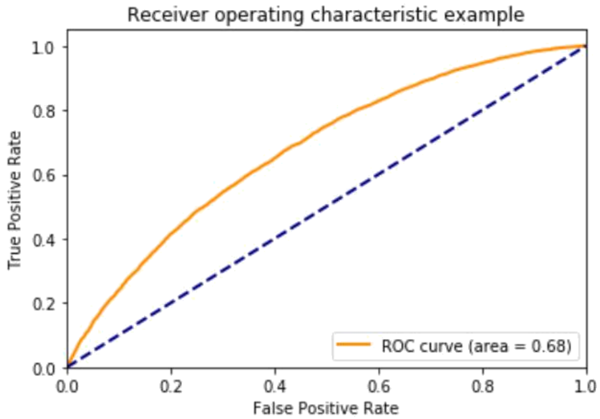

Figure 6. Decision tree.

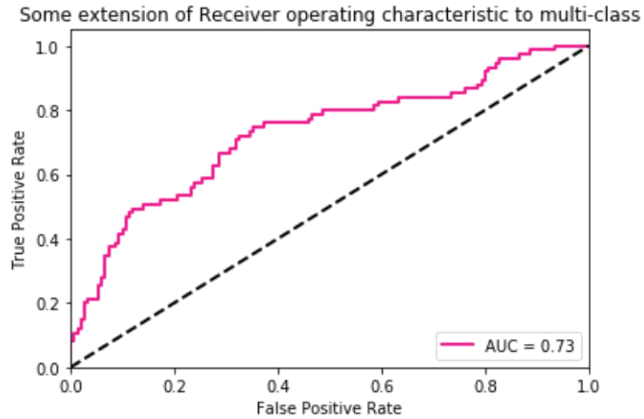

Figure 7. Random forest.

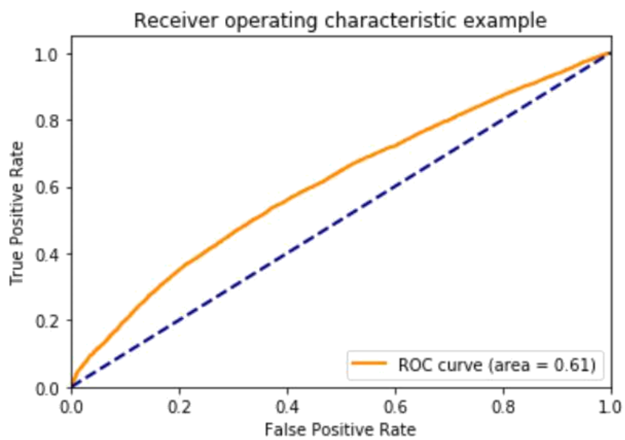

Figure 8. Simple neural network (2-Layer).

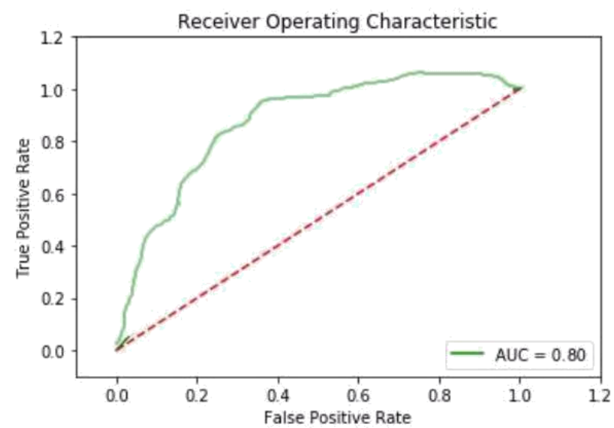

Figure 9. Recurrent neural network (2-Layer).

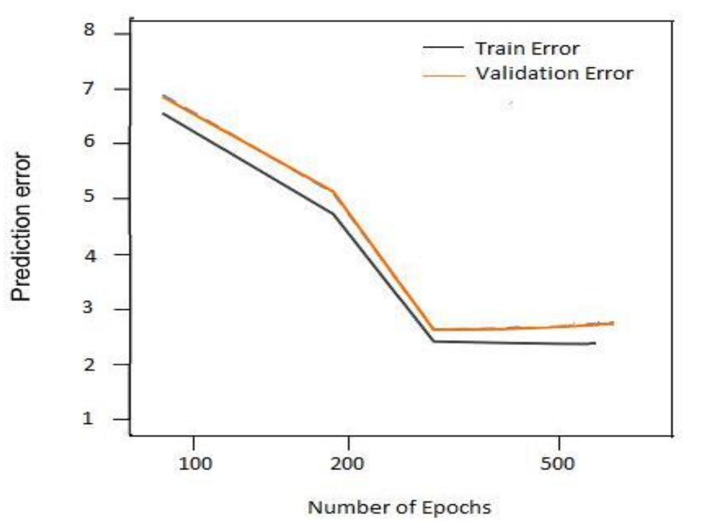

Figure 10. Validation loss and Training loss during training the data on Recurrent Neural Net.

6 CONCLUSION

Hospital Readmissions is very stressful thing for both patients and hospitals. In our research, we built a powerful model to predict the number of patients readmitted to a hospital. We compared six machine learning models to predict and concluded that Recurrent Neural Network outperformed the rest of the machine learning models in the prediction quality. We also found that the performance of RNN was remarkable on non-sequential data. This framework can be implemented in today's health system to target high risks patients, reduce rate of readmission and deliver better health care.

7 REFERENCES

[1] Strack, B., DeShazo, J. P., Gennings, C., Olmo, J. L., Ventura, S., Cios, K. J. and Clore, J. N. 2014. Impact of hba1c measurement on hospital readmission rates: analysis of 70,000 clinical data base patient records. *BioMed research inter- national*.

[2] M. P. A. Commission 2007. et al. *Report to the* Congress*: promoting greater efficiency in Medicare.* Medicare Payment Advisory Commission (MedPAC).

[3] Daras, L. C., Ingber, M. J., Carichner, J., Barch, D., Deutsch, A., Smith, L. M., Levitt, A. and Andress, J. 2017, Evaluating hospital readmission rates after dis-charge frominpatient rehabilitation. *Archives of Physical Medicine and Rehabilitation*.

[4] Freburger, J. K., Li, D., and Fraher, E. P. 2017. Community use of physical and occupational therapy after stroke and risk of hospital readmission. *Archives of Physical Medicine and Rehabilitation*.

[5] Turgeman, L. and May, J. H. 2016. A mixed-ensemble model for hospital read- mission. *Artificial intelligence in medicine* 72, 72–82.

[6] Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. 2013. *Applied logistic regression*. Vol. 398, John Wiley & Sons.

[7] Cortes, C. and Vapnik, V. 1995. Support-vector networks. *Machine learning.* 20 (3), 273–297.

[8] Cristianini, N. and Shawe-Taylor, J. 2000. An introduction to support vector machines and other kernel-based learning methods. *Cambridge university press.*

[9] Breiman, L. 2001. Random forests. *Machine learning* 45 (1), 5–32.

[10] Bengio, Y. et al. 2009. Learning deep architectures for ai, Foundations and trends R in *Machine Learning* 2 (1), 1–127.

[11] Reddy, A. 1974. A contribution to best approximation in the l2 norm. *Journal of Approximation Theory* 11 (2), 110–117.

[12] Chen, Y., Yang, J., and Qian, J. 2017. Recurrent neural network for facial land- mark detection. *Neurocomputing* 219, 26–38.

[13] Mao, J., Xu, W., Yang, Y., Wang, J., Huang, Z. and Yuille, A. 2014. Deep captioning with multimodal recurrent neural networks (m-rnn). *arXiv preprint arXiv*:1412.6632 .

[14] Larochelle, H., Bengio, Y., Louradour, J., and Lamblin, P. 2009. Exploringstrate- gies for training deep neural networks. *Journal of Machine Learning Research* 10 (Jan.2009), 1–40.

[15] Allwein, E. L., Schapire, R. E., and Singer, Y. 2000. Reducing multiclass to binary: A unifying approach for margin classifiers. *Journal of machine learning research* 1 (Dec.2000), 113–141.

[16] Fieldsend, Jonathan E., and Richard M. Everson. 2005. *Visualisation of multi-class ROC surfaces.*