Bank Failure Prediction using Deep Learning

Author
Shahid Iqbal
Registration Number
00000206562

Supervisor
Dr. Hasan Sajid

DEPARTMENT OF ROBOTICS & INTELLIGENT MACHINE ENGINEERING
SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING
NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY
ISLAMABAD

AUGUST 2021

Bank Failure Prediction using Deep Learning

Author
SHAHID IQBAL
Registration Number
00000206562

A thesis submitted in partial fulfillment of the requirements for the degree of
MS Robotics and Intelligent Machine Engineering

Thesis Supervisor:
DR. HASAN SAJID

Thesis Supervisor's Signature:______________________________________

DEPARTMENT OF ROBOTICS & INTELLIGENT MACHINE ENGINEERING
SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING
NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY, ISLAMABAD
AUGUST 2021

FORM TH-4

National University of Sciences & Technology

MASTER THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by: (Student Name & Regn No.) Shahid Iqbal (00000206562)

Titled: Bank Failure Prediction using Deep Learning be accepted in partial fulfilment of the requirements for the award of Masters in Robotics and Intelligent Machine Engineering degree. (Grade ________)

Examination Committee Members

1. Name: Dr. Jawad Khan Signature:______________

2. Name: Dr. Karam Dad Kallu Signature:______________

Supervisor's name: Dr. Hasan Sajid Signature:______________

Date:__________________

____
Head of Department

Date:______________________________

COUNTERSINGED

Date: ___________________

____
Dean/Principal

Declaration

I certify that this research work titled "*Bank Failure Prediction using Deep Learning"* is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources, has been properly acknowledged / referred.

Signature of Student
SHAHID IQBL
00000206562

Plagiarism Certificate (Turnitin Report)

This thesis has been checked for Plagiarism. Turnitin report endorsed by Supervisor is attached.

Signature of Student
SHAHID IQBAL
Registration Number
00000206562

Signature of Supervisor

THESIS ACCEPTANCE CERTIFICATE

Certified that final Copy Of MS Thesis written by

Mr. Shahid Iqbal (Registration No. 00000206562)

of Department of Robotics and Intelligent Machines Engineering (SMME) has been vetted by undersigned, found complete in all respects as per NUST Statutes / Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in this dissertation.

Signature: ______________________

Name of Supervisor: Dr. Hasan Sajid

Date: ______________________

Signature (HoD): ______________________

Date: ______________________

Signature (Dean/Principal): ______________________

Date: ______________________

Copyright Statement

Acknowledgement

I would like to thank ***Dr. Hasan Sajid*** who guided me to complete this report in time. He led me to evolve as a researcher. His professional attitude is a source of inspiration for me.

Furthermore, I would like to be grateful to my family for their endless love and support. I extend my sincere affections to all my friends and colleagues who have been a source of encouragement and gratitude and been there for every thick and thin of hostel life.

Dedicated to my family

ABSTRACT

Banking sector is principally responsible for holding financial assets/resources in any country's economy. Therefore, Bank Failure has far greater impact on the overall economy of a country compared to any other business. It can rapidly pour out to other banks and financial institutions and therefore has an avalanche effect. In order to evade fateful financial scenarios, rigorous regulations have been put in place along with technology to monitor, track and forecast critical financial parameters. Numerous statistical techniques and machine learning approaches have been widely employed for pre-emptive decision making to preclude the potential financial crisis. Banks employ domain experts, who exploit their expertise along with these tools for financial performance assessment of the financial institutions. These experts, based on performance assessment, recommend actions to prevent bank failure. Hiring domain experts exhaust substantial financial resources. Moreover, in spite of financial burden, the recommended actions do not suffice to cease bank failure most of the time because; it is very hard to generalize all the knowledge due to complex correlations of the financial parameters. The success of Artificial Intelligence (AI) across different domains attracted financial institutions to adopt much powerful AI methods to replace inefficient old methods. In an effort to employ AI for assistance in financial decision making, this research work proposes a novel deep recurrent neural network for bank failure prediction. In this work, we propose a two-layer recurrent network with Long Short-Term Memory (LSTM) cells. To validate the proposed algorithm, we collected data of 5946 banks from United States in the time span from 2004 to 2018. In total we have collected 43 financial ratios/variables over fifteen years for each of the bank. The performance of the proposed algorithm is compared against that of widely adapted SVM and Logistic Regression methods. The results vindicate the superiority of our proposed approach. The thesis work concludes with a comprehensive study of effect and role of different parameters towards bank failure.

Table of Contents

List of Figures

List of Tables

1 Introduction

Banks share a core role in financial system of an economy. Principally, they facilitate the process of wealth generation by acting as intermediary between depositors, who lend money to the banks, and borrowers, who borrow money from the banks. A bank gets into financial crisis when it becomes unable to fulfil the demands of its depositors due to wane in its assets resulting in loss of parity between its assets and liabilities. Consequently, the bank fails to keep sufficient liquidity and does not suffice to its liabilities. Therefore, it is left with two options: borrow from other stable banks or sell its assets to keep the promise with the depositors. In the latter choice, usually, assets get bid at lower price and inflict substantial loss to the bank. The bank is said to be bankrupt when it has not enough liquidity for its operation, hence it is bought out by the state financial institutions or financial regulatory body. This has generative consequences; the promulgation of bank failure infuses a panic among depositors, who, anticipating potential loss, rush to withdraw their deposits from banks. This ramp in cash withdrawals jeopardizes banks of insolvent situation. Because of this avalanche effect, the failure of banks has grave impact on the economy and is more fateful in comparison to failure of any other business. If the crisis lingers, it not only bashes the local economy but also effects the linked economies contagiously. Therefore, every state imposes stringent regulatory policies to its banks to ascertain that they may not fall to insolvent situation resulting in tumultuous situation of the financial system.

Keeping in view the criticality of the problem, extensive efforts have been made to devise numerous analytical solutions for pre-emptive prediction of bank failure as an "early warning system". Generally, such analyses require diligent execution by domain experts biannually. Moreover, the analytical process becomes extravagantly expensive due to heavy remuneration paid to such experts. Even if a bank is desirous to expend financial resources the probability of human error still prevails, after all, these experts also employ previous data to appraise the trends and prefigure future outcomes. The last few decades witnessed a whopping research in this domain, various statistical and machine learning based solutions have been concocted, each with its own advantages and disadvantages.

Existing methods typically suffer due to complex deficiencies that result in poor prediction accuracy. The selection of input features varies from one domain expert to another based on their subjective experiences. This causes lack of comprehensive information processing for result prediction. Existing methods appear naive to formulate the correlation among numerous financial parameters and their complex non-linear variations overtime. These methods overlook temporal variation of parameters, rather, they tend to map the input features to the output labels. The problem demands versatile methods that can account for temporal nature of financial parameters along with complex variations in parameter values.

To overcome the aforesaid challenges, we propose a deep Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) cells in this research thesis. Just like Artificial Neural Networks (ANN), RNNs also formulate mapping from input features to the target labels. However, at any stage recurrent neurons (basic unit of an RNN) receive both the input at current state (i.e. x_t at current time step t) and the output at previous state (i.e. y_{t-1} at previous time step t-1) contrary to artificial neurons which only take the input at current state (i.e. x_t at current time step t). These recurrent neurons, thus, exhibits a sort of memory function when connected over multiple time steps since the output of a neuron at time step t becomes dependent on all inputs from previous time steps. This attribute of recurrent neural networks makes them an ideal tool for processing of temporal data. These are even capable of learning sequences of arbitrary lengths. Complex temporal dependencies can be mapped using stacked layers of RNNs. The LSTM network (RNNs with LSTM unit cell) equipped with inherent capacity of handling temporal information makes it a prime choice for bank failure prediction. The network takes all the input features and autonomously learns to assign more weightage to important features for correct prediction. The end-to-end approach automatically compensates for any bias in selection of features. It not only captures correlation among features but also the variations over long period of time with the capability of handling variable number of time steps, a key feature that outdo previous methods. The proposed approach is validated on data of US banks that are insured by the Federal Deposit Insurance Corporation (FDIC). The data is obtained from FDIC publicly available database. The sample covers the period from 2004 to 2018 (up to three years prior to the failure of the oldest banks in the sample in 2007). The total population of failed banks in the FDIC database between the years 2007 and 2016 is 531 banks, there are no reported failed banks in 2017 and 2018, and a total of 5415 banks are non-failed in the database. The data comprise of a variety of financial features and ratios based on CAMEL rating system. To compare our proposed approach on the dataset, we compare the performance of our methodology to that of well-established machine learning methods in this domain: SVM and Logistic

Regression. The proposed approach surpasses these methods. We conclude our research by a detailed study of the paramount features imparting towards bank failure.

2 Review of Popular Traditional Approaches

An enormous number of researchers have employed, dissected, and compared various mechanisms that help in proactive prediction of bank failures. Altman pioneered the research in the field of bankruptcy in 1968 by employing discriminant analysis [1]. The last five decades have been tremendous in this field, as the research methodology have been groomed from statistical models to intelligence systems, however, the variables and the use of financial ratios for prediction of bankruptcy is almost the same.

2.1 Statistical Techniques

Over the decades, discriminant Analysis (DA) has been one of the most acceptable and reliable statistical technique in determining bank failures [2-3]). DA has been classified into three subcategories i.e., Linear, Multivariate and Quadratic. DA is considered to be supreme in analyzing cross-sectional data. Whereas, time series data is frequently analyzed by harnessing hazard or duration analysis models. To carry out DA, regressors should have normal distribution. In case, regressors are not normally distributed, maximum likelihood method of Logit is employed. Logit is an abridged form of Logistic Regression, which is a statistical analysis technique that makes use of logistic function for modelling of binary dependent variable.

West suggests that factor analysis combined with Logit estimation gives a valuable insight for assessment of banks' working conditions [4]. West findings also vindicates that the factors marked as significant variables in determining banks' operating conditions have close resemblance with CAMELS ratings. CAMELS rating system was originally developed in US for the classification of bank's overall condition. The components of a bank that are addressed by this rating system are (C)apital adequacy, (A)ssets, (M)anagement capability, (E)arnings, (L)iquidity and (S)ensitivity.

Early Warning Systems (EWS) have been widely opted by central banks to keep an eye on bank risks, on the other hand, "Federal Deposit Insurance Corporation Improvement Act of 1991" made it statutory for banks to conduct annual or eighteen-month examination cycle. The regulators use CAMELS rating framework for assessment of each bank's wellbeing and soundness. Davis & Karim exploited statistical intelligence techniques for assessment of potential bank crises [5]. Their study premises a comparison of logistic regression (Logit) and signal extraction EWS. They concluded that the choice of estimation models sorely impacts the performance of variables and eventually the crises predictability. Logit models display a better predictability performance in global EWS whereas signal extraction disposes itself as a dependable forecaster in country specific EWS. In another independent study, Davis & Karim essayed Logit and binomial tree based EWS that helped in prediction of bank failures in US and UK [6]. The results are opinionated that the Logit performance exceeds than that of rest of the techniques.

An Integrated Early Warning System (IEWS), based upon EWS, was purported by Canbas [7]. IEWS unifies Logit, DA, Probit and Principal Component Analysis (PCA) approaches to anticipate bank failures. The system initially employs PCA for selection of paramount financial components in defining the condition of banks. After the selection of dominant financial parameters DA, Logit and Probit regression models are used for financial condition assessment. IEWS exhibits a better performance w.r.t predictability compared to most single models used in available literature.

2.2 Intelligent Approaches

Intelligence modelling techniques have been extensively applied in operational research to forecast bank failures and financial crises. One of the most widely employed intelligence techniques is Artificial Neural Networks (ANNs). ANN models make use of mathematical and algorithmic kernels that imitate biological neural networks of the human nervous system. Celik & Karatepe utilized artificial neural networks to bode crises [8]. On the other hand, Alam et al. utilized fuzzy clustering and self-organizing neural network to identify failed banks [9].

Boyacioglu et al., in his research on bank failures in Turkey, contested various NN, Support Vector Machine (SVM), Multivariate Discriminant Analysis, Cluster Analysis and Logit regression analysis in the perspective of CAMELS rating [10]. The results vindicated that Multivariate Discriminant Analysis and Logit regression analysis perform better in failure prediction among all others.

Tam employed a multilayer NN, known as Back-Propagation Neural Network (BPNN) model to successfully predict Texas bank failures almost one to two years prior the collapses [11]. He used CAMELS variables in his investigations and reasoned that BPNN outmatched K-nearest neighbor, DA and Logit technique, in accurate prediction of bank failures. Tam & Kiang, in another study, exploited linear discriminant analysis (LDA), Logit, K-Nearest Neighbour, Interactive Dichotomizer 3 (ID3), feedforward NN and BPNN for pre-emptive bank failure prediction [12]. Out of all the models used, BPNN surpassed rest of the models for one-year prior samples, whereas LDA got over the rest for two years prior samples. However, BPNN surmounted all, in both one and two-year prior samples for holdout samples and in jackknife method. They conclusively state that NN outdoes DA method.

In his study for prediction of bank failure using Logit and BPNN models, Bell argues that neither Logit nor BPNN model has any superiority to each other in predictability but, when it comes to make complex decisions BPNN exhibits its superiority to other contending model [13]. Swicegood & Clark, in their investigation, conclude that BPNN better perform in recognizing underperforming banks [14]. Their study subjects to comparison of DA, BPNN and human judgment in bank failure prediction.

2.3 Logit and Trait Recognition (TR) Approaches

Trait Recognition models formulate another approach for prediction of bank collapses. These models are devised from different distribution sections for each variable based on its interactions with one or more variables in the distribution sections. Two sets of safe and unsafe traits are used as discriminators; bank failure is anticipated by classifying each bank under one of these discriminators. Trait recognition can discern complex correlation of variables. The distinguishing feature of this approach is that it establishes a cut off point for each variable in such a way that all failed banks are placed underneath this threshold point and survived banks reside over it. Kolari et al. applied EWS based Logistic Regression and Trait Recognition methods on a large set of US banks [15]. The Logit model predicted bank failures one and two year prior to their shutdown with an accuracy of 96%. On the other hand, TR model, outperformed Logit with an accuracy of 100%.

Lanine & Vander Vennet also employed Logit and TR approaches for prediction of bank failures in Russia [16]. They contested the two models based on prediction accuracy. Their results upheld the superiority of TR approach over Logit in holdout and original samples. They concluded that liquidity, asset quality and capital adequacy played a major role in determination of bank failures.

There are some machine learning techniques such as Decision Trees (DT), which apply "recursive partitioning algorithm" to formulate patterns on a given data set. Algorithms such as classification and regression trees (CART) can also be employed effectively to cater for problems that require accurate prediction [17-18]) i.e. health problems and financial performance analysis.

2.4 Support Vector Machine

Support Vector Machine (SVM) technique is based on the principle of Structural Risk Minimization (SRM). SRM principle finds its basis in computational learning theory introduced by Vapnik [19]. Input data comprising of different classes are arranged in multidimensional space. SVM model, employing a specialized linear model, outputs an ideal hyper plane that achieves extreme division between the two classes. Many a researcher have exploited this technique for prediction of bank failure including Boyacioglu et al., Vapnik et al., Chen & Shih et al. and Huang et al. [10, 19-21]. Shin et al. argues that SVM technique surpasses BPNN in accurate prediction of bankruptcy of financial institutions [22].

2.5 Hybrid Approaches

Kosmidou & Zopounidis devised a multi-criteria decision technique; UTilites Additives DIScrim-inants (UTADIS) for prediction of bank failure [23]. UTADIS performs well with "ordinal classification problem" and is robust to statistical problems because the additive utility function is not performed through statistical methods rather mathematical linear programming technique is employed for it. UTADIS outperforms traditional multivariate data analysis techniques.

Multicriteria Decision Aid (MCDA) technique is used for determination of credit ratings and bank soundness. This method exhibits a superior performance in comparison to conventional multiple discriminant analysis. Gaganis et al. made use of MCDA model using the UTADIS method for categorization of banks based on their soundness [24]. The model employed cross validation procedure. The results depict that the most significant criteria for categorization of bank soundness is based on capitalization, asset quality and banks' operating market. UTADIS exhibits highest classification accuracy in comparison to DA and logit.

A multiple criteria decision-making framework named Analytic Network Process (ANP) has also been employed for prediction of financial crises. Niemira & Saaty, based on their investigations, premised that ANP framework is more flexible and comprehensive in contrast to other traditional models and thus serves as a good choice to forecast crises [25]. Owing to its structural construction, ANP framework also minimizes judgmental forecast errors.

There are other hybrid methodologies used over time to time. Ravi & Pramodh, concludes that hybrid models which combine Principal Component Neural Network (PCNN) with failure prediction models, outperform other traditional classifiers available in literature [26].

2.6 Other Approaches

A mathematical method termed as Rough Set technique models imprecise and partial data; a concept first described by Pawlak [27]. It applies estimation of rough objective into predefined classes under examination (for more details see Greco et al. [28]). Ahn et al. combined rough set theory with artificial neural networks for prediction of failures [29]. This hybrid model performed exceedingly well against discriminant analysis and neural network models.

Case base Reasoning (CBR) is another approach that offers a good accuracy in predictions. It makes use of past experiences to predict failure. Decision making in CBR comprises of four steps. First, retrieval of similar cases. Second, use of retrieved cases to solve the problems. Third, if possible, review and revise proposed solution. Fourth, keep new arrangement as part of new case.

Nearest Neighbor technique classifies an object, in our case a bank, based on its similarity index with other objects (banks). Banks are designated to a class, i.e., survived or failed, based on most common class amongst its K nearest neighbors.

3 DATA SET

The data set was composed of data from 5946 US banks that were insured by the Federal Deposit Insurance Corporation (FDIC). The data is obtained from FDIC publicly available database. The sample covers the period from 2004 to 2018 (up to three years prior to the failure of the oldest banks in the sample in 2007). The total population of failed banks in the FDIC database between the years 2007 and 2016 is 531 banks, there are no reported failed banks in 2017 and 2018, and a total of 5415 banks are non-failed in the database.

United States has the biggest and most advanced economy in the world. Its share in the world's total GDP is 24.41% with a total GDP of $ 21.43 trillion. The banking system of US is one of the advanced banking systems in the world. Therefore, the data set becomes a good representative for bank failure study. The dataset contains a total of 84676 samples with each sample formed by 43 different financial parameters and financial ratios. The parameters/ratios have been categorized using the CAMELS rating system, originally known as the Uniform Financial Institutions Rating System (UFIRS) (Council, 1996).

Table 3-1, Table 3-2, Table 3-3, Table 3-4, Table 3-5 present these parameters and ratios for each of the categories: capital, asset, management, earnings, and liquidity respectively. There is no financial parameter available in the data set under the category of sensitivity.

3.1 CAMELS Rating System

Let's have a brief discussion on each category of the CAMELS rating system. *Capital adequacy* indicates capital condition of a bank. It also shows resilience of a bank to cater for the losses and preclude its operations from ceasing. *Asset quality* in the CAMELS rating system is a measure of the efficiency of bank's investment policies and practices. It is determined by rating the risk factors a bank may experience on its assets. *Management quality* in the CAMELS rating system is a measure of potential and effectiveness of top-level management which owe successful operations of a bank. It also accounts their competency to assess, adapt and retort to the market trends. *Earnings* category defines a bank's long-term viability. It is depicted as the return on assets ratio. It takes in the income of a bank from all available sources i.e., operations, and other non-traditional sources. *Liquidity* is the potential of a bank to transform its assets to cash. It is measured as the ratio of cash holdings of a bank to its total assets. *Sensitivity* is described as delicacy and fragileness of a bank against the market risks. It reflects the degree to which earnings are affected by adverse market changes such as foreign exchange rates, commodities prices etc.

3.2 Challenges with the dataset

The dataset offers two main challenges: imbalanced data and variable sequence length. Imbalanced data means that data do not contain equal number of samples for failed and survived banks. The data contains 91.07% banks that survived against only 8.93% banks that collapsed. Variable sequence length depicts that all the samples were not 15 units long (from 2004-2018). For example, some of the banks failed earlier than 2018 and some banks failed after 2007.

Table 3-1: *List of capital features*

Global S. No.	Category	S. No.	Feature Description
1	Capital	1	Retained earnings to average equity
2		2	Equity capital to assets
3		3	Core capital (leverage) ratio
4		4	Tier 1 risk-based capital ratio
5		5	Total risk-based capital ratio
6		6	Pretax income to equity
7		7	Liabilities to Equity
8		8	(income + equity) to total assets
9		9	Total equity to gross loans
10		10	(Share holders' equity + total income)/ (deposits + other borrowed funds)

Table 3-2: *List of assets features*

Global S. No.	Category	S. No.	Feature Description
11	Assets	1	Noninterest income to average assets
12		2	Loan and lease loss provision to assets

13		3	Net charge-offs to loans
14		4	Credit loss provision to net charge-offs
15		5	Earnings coverage of net charge-offs
16		6	Earning assets to total assets ratio
17		7	Loss allowance to loans
18		8	Loan loss allowance to noncurrent loans
19		9	Noncurrent assets plus other real estate owned to assets
20		10	Noncurrent loans to loans
21		11	Net loans and leases to total assets
22		12	Total domestic deposits to total assets
23		13	liabilities to assets
24		14	Total expenses to total assets

Table 3-3: *List of management features*

Global S. No.	Category	S. No.	Feature Description
25	Management	1	Assets per employee

Table 3-4: *List of earnings features*

Global S. No.	Category	S. No.	Feature Description
26	Earnings	1	Yield on Earnings Assets
27		2	Cost of funding earnings assets
28		3	Net interest margin
29		4	Noninterest expense to average assets
30		5	ROA
31		6	Pretax ROA

32		7	ROE
33		8	non-interest expense to interest and non-interest income
34		9	Cash dividends to net income
35		10	Retained earnings to total assets
36		11	Net income growth rate
37		12	Income before extraordinary items to total assets
38		13	Interest expense to total expense

Table 3-5: *List of liquidity features*

Global S. No.	Category	S. No.	Feature Description
39	Liquidity	1	Net loans and leases to deposits
40		2	Net loans and leases to core deposits
41		3	Cash to Total assets
42		4	Gross loans to total deposits
43		5	Cash to Total liabilities

4 PROPOSED APPROACH

Before diving into the details of our proposed approach, let's have a brief discussion on the theory of Recurrent Neural Networks (RNNs), their differences to widespread ANNs and their suitability to process financial data. Generally, financial parameters and features are non-linear, extremely correlated and possess temporal traits. The temporal nature of financial parameters suggests that their correlation does not map a single time instance to an exact single time instance rather it exists over multiple time steps. Fluctuation in one parameter can become a source of turbulence in other parameters and therefore vexes the gross output or contribution towards bank's survivability. For example, if the liquidity ratios undergo a positive change, it will seem that bank can sleekly operate but, it may also imply that the bank is refraining from capital investment. Thus, in the longer run, the earnings of the bank will suffer a negative impact. And if earnings of the bank decrease, the result will be a loss to the bank or a relatively lower profit in its capital investments. The loss will superimpose a negative impact on its capital and bank will face a reduction in its capital. Thus, apparently an auspicious change in liquidity ratios induces an ominous consequence on the overall bank performance.

4.1 Artificial Neural Networks

Artificial Neural Networks are a result of endeavors of the inspiration from human neurological system that consists of inter-connected neurons in a mesh like configuration. ANNs consist of connected nodes called artificial neurons analogous to human neurons, the threads connecting these artificial neurons are analogous to synapses. These connecting threads are assigned weights. The network attempts to adjust these weights to accomplish the task for a given problem. A typical ANN architecture has been shown in Figure 4-1.

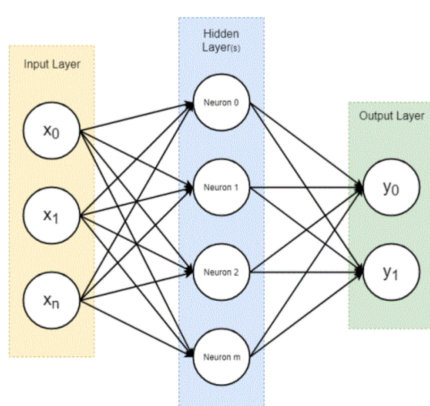

Figure 4-1: *Typical ANN Architecture*

The neural networks are built in layered structure. These layers belong to three different categories: input layer, hidden layer(s) and the output layer. ANNs possess inherent traits to map non-linear data as they have the capability of learning complex inter-feature dependencies. Unfortunately, ANNs cannot discern temporal property of the data. Therefore, ANN models do not have the capacity to learn any temporal feature variations and dependencies.

4.2 Recurrent Neural Networks

The inability of ANNs to tackle temporal feature of data led to a newer class of neural networks. This newer class is termed as Recurrent Neural Networks (RNNs) that account for temporal feature along with possessing the advantages of ANNs. An RNN is almost like an ANN, however, in contrast to ANNs, where activations flow only in forward direction, RNN also possesses connections that point in backward direction. Figure 4-2 depicts a recurrent neuron and multiple neurons inter-connected temporally.

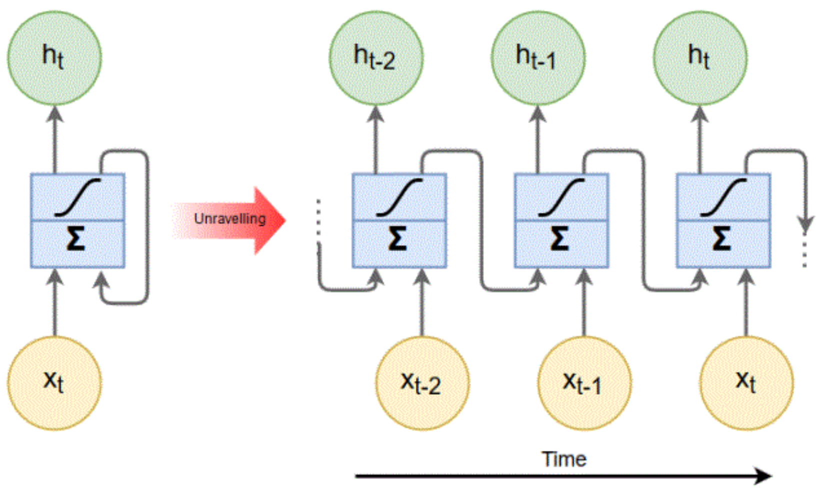

Figure 4-2: *Single RNN cell (left), unraveled single RNN cell*

A recurrent neuron has been depicted in Figure 4-2 (left). The recurrent neuron gets an input and produces an output, this output is also fed back to the recurrent neuron to take part in the assessment of next output. Figure 4-2 (right) depicts a situation when multiple inter-connected neurons are unrolled through time, the figure clearly shows that at each time step every neuron takes in two inputs: an input at the current time step and another input from output at the previous time step. Therefore, each recurrent neuron has separate sets of weights for these two inputs, thereby, possessing a form of memory. Just like ANNs, multiple layers of RNNs can be stacked with single or multiple recurrent neurons as shown in Figure 4-3.

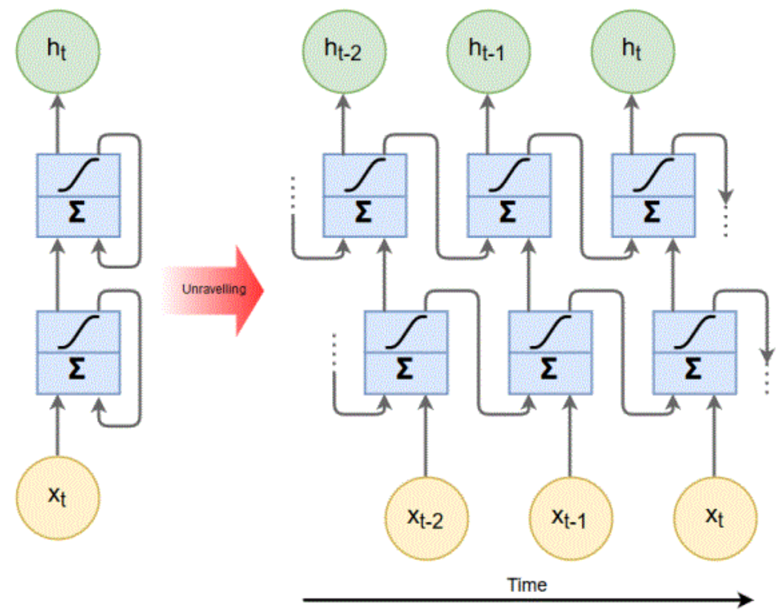

Figure 4-3: *Deep recurrent neural network*

4.3 RNN Training Challenge

Just like ANNs suffer vanishing/exploding gradient problem during training process, RNNs are also unable to preclude this ominous problem. A brief discussion on this challenge is presented here.

4.3.1 Vanishing/Exploding Gradient Problem

Unfortunately, the vanilla RNNs are no exception to the challenge of vanishing/exploding gradient problem, especially when networks get deep [30]. In this problem, as the algorithm traverses down to the lower layers, gradients start getting smaller and smaller. Consequently, the Gradient Descent update becomes handicapped in tuning the lower layer connection weights, and thus, training never converges to an optimal solution. This problem is termed as the vanishing gradients problem. In some cases, the opposite happens: the gradients start getting bigger and bigger, so many layers get freaky due to large weight updates and the algorithm diverges.

LSTMs and GRUs are renowned versions of basic RNN cell that were proposed by Hochreiter & Schmidhuber, and Chung respectively [31-32]. Their objective was to counter gradient vanishing problem through rerouting gradient by supplying an alternat path. We have used LSTM version of RNNs in our proposed approach. Furthermore, LSTMs provide the advantage of rapid convergence during training, can automatically identify important features and retain long-term dependencies in the data.

4.4 Proposed Network Architecture

We propose a 2-layer deep recurrent neural network made up of LSTM cells. as shown in Figure 4-4. There are 64 constituent recurrent neurons in each cell. The data set has been normalized using Equation 4.1 before it is fed to the network. The objective of normalization is to ensure that no parameter dwarfs the other only because of its magnitude.

$$x_{norm} = \frac{x - Min(x)}{Max(x) - Min(x)} \qquad \text{Equation 4.1}$$

Where x_{norm} is the normalized value of feature x, $Min(x)$ is the minimum value of feature x and $Max(x)$ is the maximum value of feature x. The network takes a three-dimensional block of input with size $n \times \delta \times \beta \mid n, \beta, \delta \in N$ where n is the batch size, δ is the number of time steps and β is the number of financial features. The network is formulated so as to take variable sequence length input, a trait associated with the data set. The network processes the input data and categorizes the bank in one of two classes i.e., failed or survived.

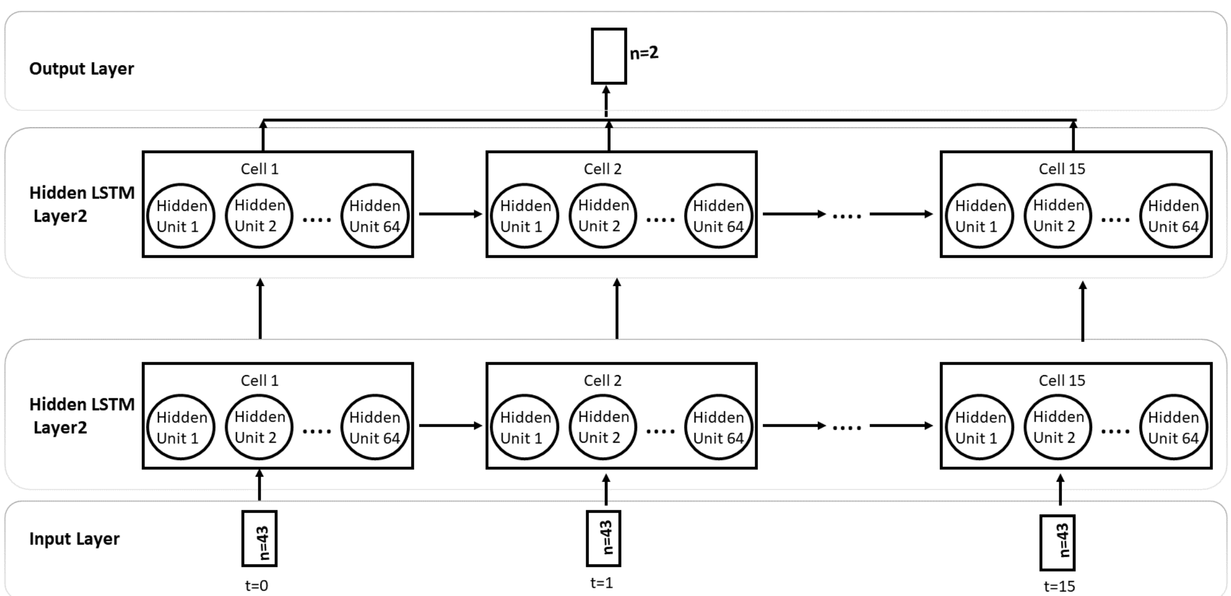

Figure 4-4: *Proposed LSTM model*

4.5 Internal Working of LSTMs

The constituent neurons of an LSTM network are also termed as LSTM cells. Each LSTM cell has three internal gates: *forget gate, input gate and output gate*. These gates serve as principal controller for flow of information. Each gate can get a value between 0 and 1, where 0 implies that the input is blocked and 1 implies that the input is allowed completely to pass through the gate. Figure 4-5 shows an LSTM cell.

Internally the state of LSTM cell is split in two vectors: $h_{(t)}$ and $c_{(t)}$ ("c" stands for "cell"). $h_{(t)}$ can be regarded as the short-term state and $c_{(t)}$ as the long-term state. As the long-term state $c_{(t-1)}$ travels through the network from left to right, it goes into the forget gate first, the gate drops some memories and adds some new memories as well. This addition is done through addition operation on selected memories by input gate. The result $c_{(t)}$, without any further metamorphism, is sent out. Thus, memories are added and dropped at each timestep. After the addition operation is done, a copy of the long-term state is retained and then passed through the tanh function, and in the end, the output gate filters output $y_{(t)}$. The short-term state $h_{(t)}$ is equal to the cell's output $y_{(t)}$ for the same time step. Thus, an LSTM cell is capable to recognize an important input and can store it in the long-term state as per the need.

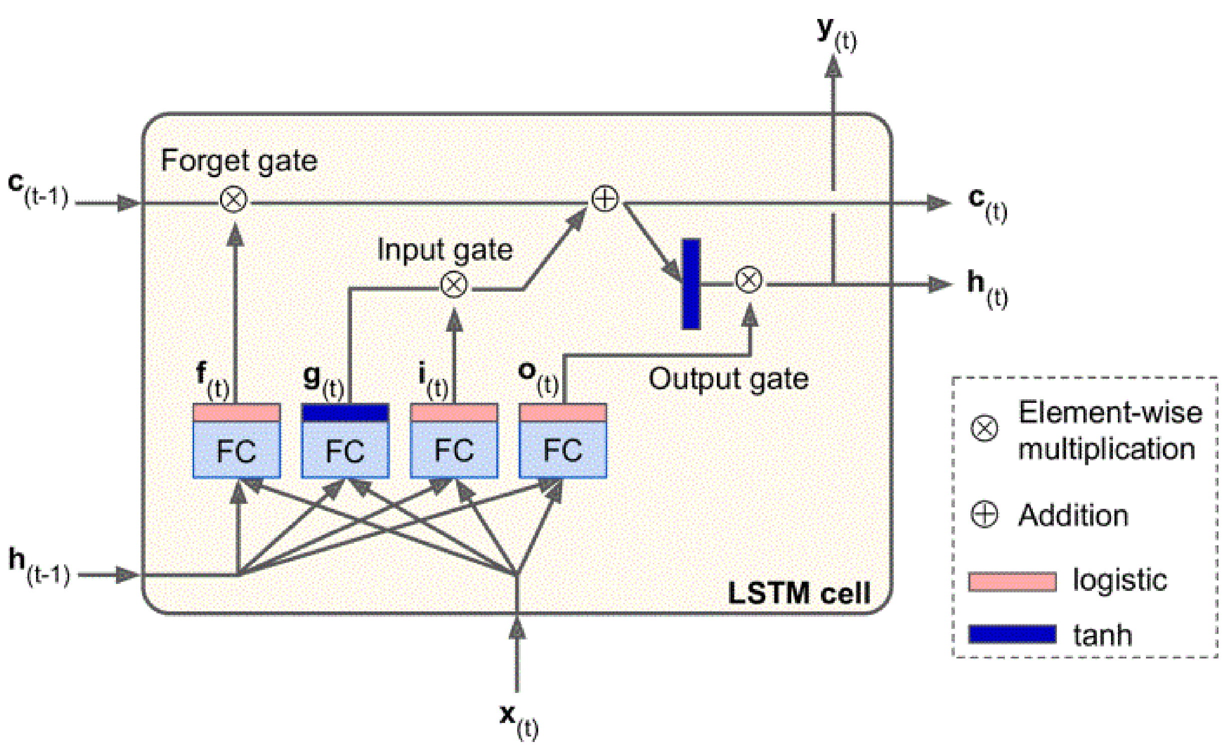

Figure 4-5: *LSTM cell architecture*

Equations 4.2 to 4.7 presents the highlights to compute the long-term state, short-term state, and output of the LSTM cell.

$$i_{(t)} = \sigma\left(w_{xi}^{T} \cdot x_{(\tau)} + w_{hi}^{T} \cdot h_{(t-1)} + b_i\right) \qquad \text{Equation 4.2}$$

$$f_{(t)} = \sigma\left(W_{xf}^{T} \,.\, x_{(t)} + W_{hf}^{T} \,.\, h_{(t-1)} + b_f\right) \qquad \text{Equation 4.3}$$

$$o_{(t)} = \sigma\left(W_{xo}^{T} \,.\, x_{(t)} + W_{ho}^{T} . h_{(t-1)} + b_o\right) \qquad \text{Equation 4.4}$$

$$g_{(t)} = tanh\left(W_{xg}^{T} \,.\, x_{(t)} + W_{hg}^{T} . h_{(t-1)} + b_g\right) \qquad \text{Equation 4.5}$$

$$c_{(t)} = f_{(t)} \otimes c_{(t-1)} + i_{(t)} \otimes \mathrm{g}_{(\mathrm{t})} \qquad \text{Equation 4.6}$$

$$y_{(t)} = h_{(t)} = o_{(t)} \otimes \tanh\left(\mathrm{c}_{(\mathrm{t})}\right) \qquad \text{Equation 4.2}$$

The data set has non-linear financial features that are highly inter-correlated along with their temporal nature. LSTMs contain non-linear activation functions e.g., ReLU and Sigmoid, thereby, possessing the capacity to tackle non-linearities in the financial data. The complex inter feature dependencies have been tackled by the proposed deep layered structure. The recurrent links on the neurons exhibit a form of memory where each neuron can remember its previous state. These properties of LSTMs make them a good choice for processing non-linear, and temporal financial data at hand.

5 EXPERIMENTS AND RESULTS

In this chapter we present the results associated with our implementation of baseline and proposed methods. The comparison among the performance of these methods has been established in this section. At the end, a detailed study to assess the importance of different financial features towards bank failure has been shared.

5.1 Performance Metrics

Performance metrics provide a common ground to evaluate the effectiveness of different techniques on a given problem. We have chosen accuracy, precision, recall and f1-Score as the metrics for comparison of performance among the baseline and our proposed models. We have formulated a confusion matrix to evaluate the performance metrics, Table 5-1.

Table 5-1: *Confusion matrix*

Class		**Prediction**	
		Survived	**Failed**
Actual	**Survived**	T_N	F_P
	Failed	F_N	T_P

The description of confusion matrix is as follows:

T_P = True positive (both actual and predicted classes are failed),

T_N = True negative (both actual and predicted classes are survived),

F_P = False positive (actual class is survived but predicted class is failed)

and

F_N = False negative (actual class is failed but predicted class is survived).

5.1.1 Accuracy

Accuracy is defined as, "the ratio of correct predictions to the total number of predictions." Equation 5.1 serves as formula to calculate accuracy.

$$Accuracy = \frac{T_P+T_N}{T_P+F_P+T_N+F_N} \quad \text{Equation 5.1}$$

5.1.2 Precision

Precision is defined as, “the ratio of true positive predictions to the total number of positive predictions.” Equation 5.2 serves as formula to calculate accuracy.

$$Precision = \frac{T_P}{T_P+F_P} \quad \text{Equation 5.2}$$

5.1.3 Recall

Recall is defined as, “the ratio of true positive predictions to the predictions that should have been positive if the algorithm was absolutely perfect.” Equation 5.3 serves as formula to calculate accuracy.

$$Recall = \frac{T_P}{T_P+F_N} \quad \text{Equation 5.3}$$

5.1.4 F-1 Score

Precision and Recall are good measures of performance. However, in statistical analysis the harmonic mean of precision and recall is considered most reliable metric. The harmonic mean of precision and recall is termed as F-Measure or F-1 Score. Equation 5.4 serves as formula to calculate accuracy.

$$F\,1\,Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad \text{Equation 5.3}$$

5.2 Baseline Implementations

We have selected two of the most popular prediction methods: SVM and Logistic regression as baseline methods.

5.3 Support Vector Machines

Since, the prediction of bank failure is a binary classification problem, therefore SVM becomes a natural choice for problem at hand. The features vector with dimension $\beta \times 43 \mid \beta \in N$ is used to train SVM, where β is the batch size hyper-parameter. The problem has been devised as a two class soft-margin Support Vector Classification with regularization parameter *C* and the radial basis function (RBF) [33]. The RBF kernel is defined in Equation 5.5.

$$K(x_i, x_j) = e^{-\gamma\left|\left|x_i - x_j\right|\right|^2} \quad \text{Equation 5.5}$$

The values of regularization parameter *C* and shape parameter γ have been set as their best combination that yields the best overall performance with a 10-fold cross-validation over training data [34]. The experimentation for C and γ suggested their values to be 1 and 0.25. Therefore, the results shown in Figure 5-1, have been deduced at C = 1 and γ = 0.25.

5.4 Logistic Regression

Logistic Regression (LR) is another baseline method that has been employed in our study. LR maps one or more independent variables, in our case 43 financial features, to a dependent variable, in our case probability of bank failure. The function of logistic regression has been given as Equation 5.6.

$$p(x) = \frac{1}{1+e^{-(\theta_0+\theta_1 x_1+\theta_2 x_2+\cdots+\theta_n x_n)}} \qquad \text{Equation 5.6}$$

Here $p(x)$ is the probability of bank failure, whereas x_1, x_2 ... x_n are the input financial features and θ_0, θ_1 ... θ_n are the learning coefficients associated with the input financial features. Experimentation has been done with different hyperparameters and 10-fold cross validation approach has been used over training data [34]. The best results have been shown in Figure 5-1.

5.5 LSTM Model Implementation

All the experimentation has been performed on Google Colaboratory. We have implemented and trained 2-layer deep recurrent neural network based on LSTM cells using Google's Tensorflow 2. The network has been devised to tackle variable sequence length input. The best results have been reported in Figure 5-1. The network parameters are as follows:

- Number of input features: 43
- Output classes: 2
- Maximum sequence length: 15
- Number of units in cell: 64
- Network layers: 2
- Optimizer: Adam optimizer
- Learning rate: 0.001
- Loss function: Cross entropy loss
- Number of epochs: 10
- Batch size: 2
- Train set: 3546 sequences
- Test set: 1200 sequences
- Cross validation set: 1200 sequences

The network has been trained on train data set and the performance metrics have been evaluated on the test set.

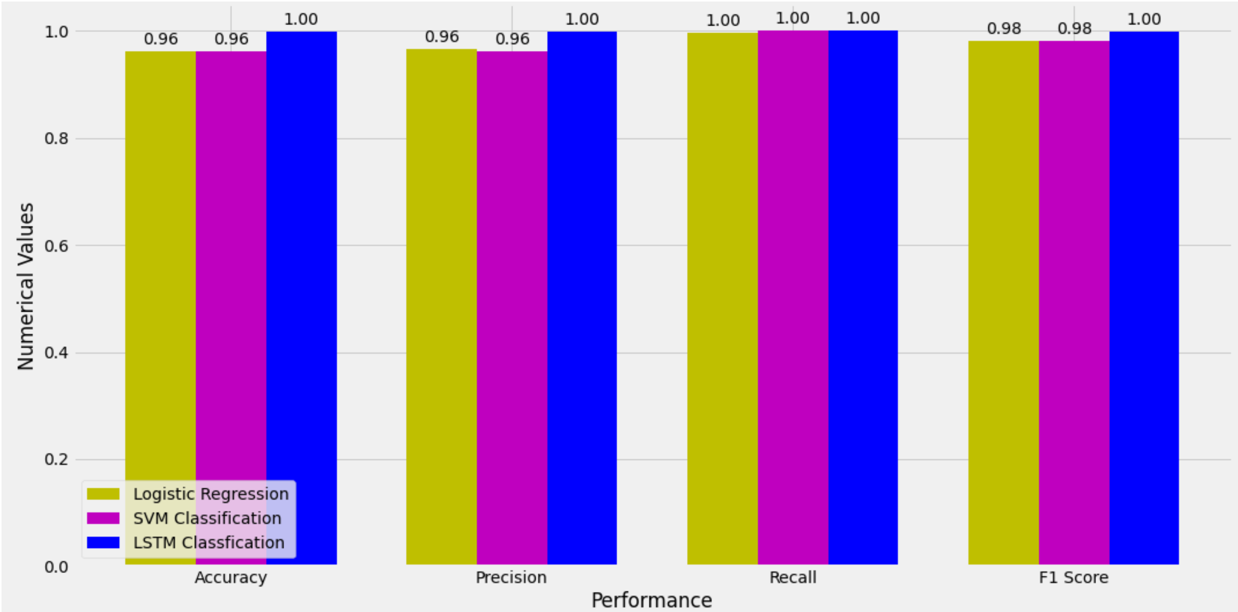

Figure 5-1: *LSTM performance comparison with baseline implementations*

5.6 Ablation Study

Ablation study is the analysis of the performance of a machine learning algorithm in which a part of data is kept hidden from it. This analysis is performed to ensure that the algorithm is not facing any over-fitting problem. Additionally, it offers aid for the identification of key features in a complex data. In this study three different techniques have been employed to test our proposed algorithm.

5.7 Experiments with Layers

Four different LSTM networks with one, two, three and four hidden layers have been formulated and trained. Figure 5-2 shows the performance graph of these networks. Based on F1 Score and number of network parameters, networks with one layer and two layers perform optimally regarding acceptable score and network complexity.

5.8 Experiments with Batch Size

Five different two-layered LSTM networks have been formulated and trained on various input batch sizes. Figure 5-3 shows the performance graph of these networks. Based on F1 Score, the batch size experiment suggests a best batch size of 8 examples per batch.

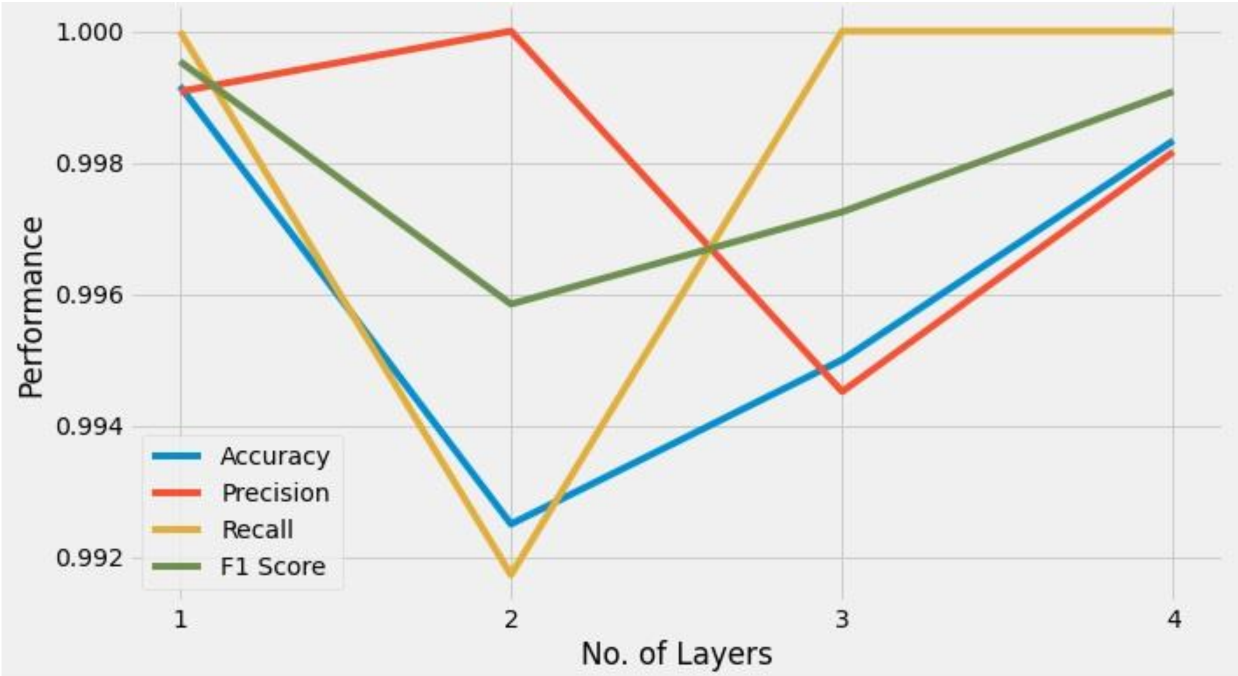

Figure 5-2: *Experiment with number of hidden layers*

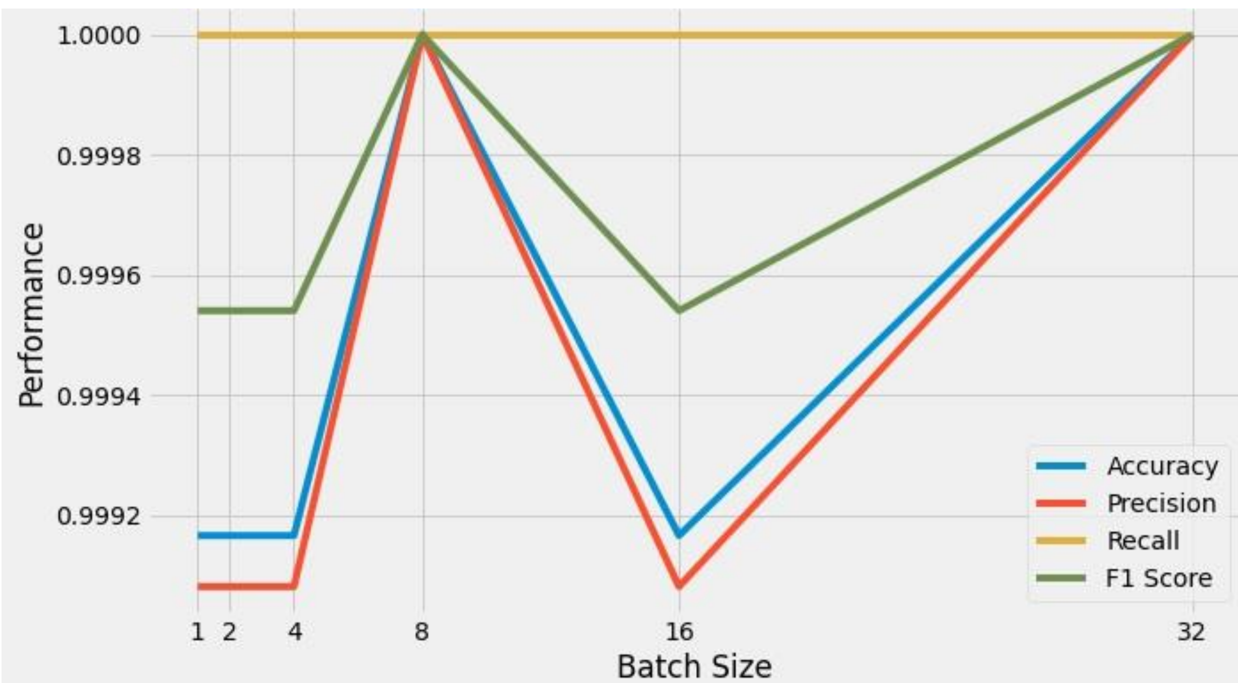

Figure 5-3: *Experiment with batch size*

5.9 Feature Importance Experiment

It is impossible to visualize internal weights of an RNN network due to the masking of complex feature correlation and temporal variations. Therefore, one cannot directly deduce the important features that influence the results. To evaluate the importance of various categories of CAMELS rating system, we have devised a permutation technique. In this technique, all the financial parameters belonging to a category are randomized keeping the other parameters retained in original state and the performance of the network is evaluated. This activity is repeated for each CAMELS category. The decrease in performance by this randomization can be implied for the important features that have the most influence on the results. Figure 5-4 shows decrease in performance for each category. Keeping in view the number of parameters in each category and the % decrease in performance we can premise the importance of these categories as below:

Management > Liquidity > Capital > Earnings > Assets

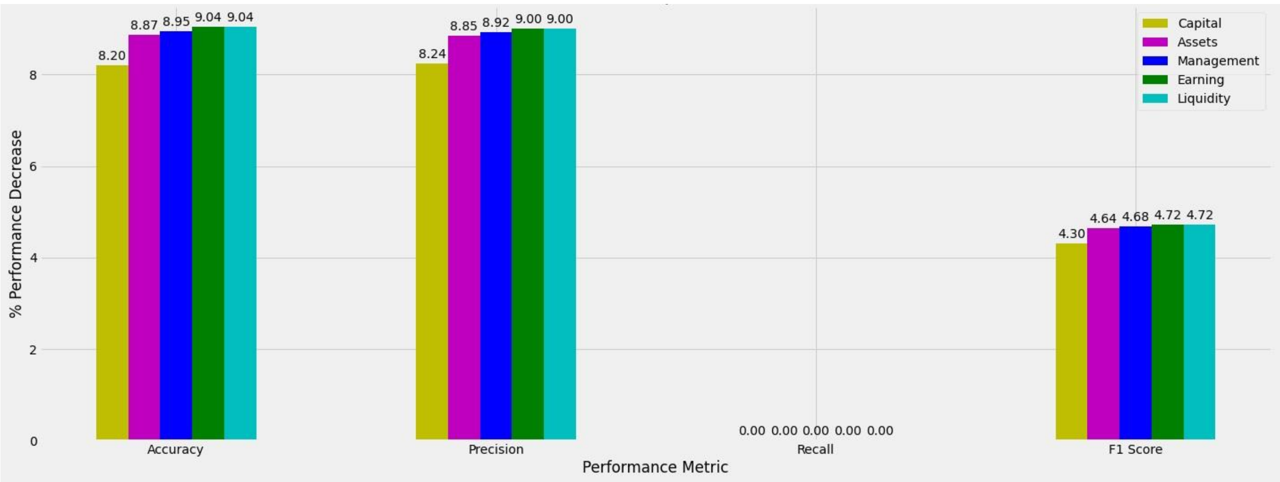

Figure 5-4: *Feature importance*

6 Conclusion and Future work

The results depict that the LSTM network can substantially surpass Logistic Regression and SVM. The cardinal ground behind these favorable results for LSTM network is that Logistic Regression and SVM do not address the temporal nature of the data whereas LSTM network retains the temporal variation of the data in its memory across the years.

For future work it is recommended that Principal Component Analysis (PCA) may be employed for feature selection and thus reduced but important features may be used for training of the LSTM model.

REFERENCES

[1] Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, *23*(4), 589-609.

[2] Karels, G. V., & Prakash, A. J. (1987). Multivariate normality and forecasting of business bankruptcy. *Journal of Business Finance & Accounting*, *14*(4), 573-593.

[3] Haslem, J. A., Scheraga, C. A., & Bedingfield, J. P. (1992). An analysis of the foreign and domestic balance sheet strategies of the US banks and their association to profitability performance. *MIR: Management International Review*, 55-75.

[4] West, R. C. (1985). A factor-analytic approach to bank condition. *Journal of Banking & Finance*, *9*(2), 253-266.

[5] Davis, E. P., & Karim, D. (2008). Comparing early warning systems for banking crises. *Journal of Financial stability*, *4*(2), 89-120.

[6] Davis, E. P., & Karim, D. (2008). Could early warning systems have helped to predict the sub-prime crisis?. *National Institute Economic Review*, *206*, 35-47.

[7] Canbas, S., Cabuk, A., & Kilic, S. B. (2005). Prediction of commercial bank failure via multivariate statistical analysis of financial structures: The Turkish case. *European Journal of Operational Research*, *166*(2), 528-546.

[8] Celik, A. E., & Karatepe, Y. (2007). Evaluating and forecasting banking crises through neural network models: An application for Turkish banking sector. *Expert systems with Applications*, *33*(4), 809-815.

[9] Alam, P., Booth, D., Lee, K., & Thordarson, T. (2000). The use of fuzzy clustering algorithm and self-organizing neural networks for identifying potentially failing banks: an experimental study. *Expert Systems with Applications*, *18*(3), 185-199.

[10] Boyacioglu, M. A., Kara, Y., & Baykan, Ö. K. (2009). Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey. *Expert Systems with Applications*, *36*(2), 3355-3366.

[11] Tam, K. Y. (1991). Neural network models and the prediction of bank bankruptcy. *Omega*, *19*(5), 429-445.

[12] Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: the case of bank failure predictions. *Management science*, *38*(7), 926-947.

[13] Bell, T. B. (1997). Neural nets or the logit model? A comparison of each model's ability to predict commercial bank failures. *Intelligent Systems in Accounting, Finance & Management*, *6*(3), 249-264.

[14] Swicegood, P., & Clark, J. A. (2001). Off-site monitoring systems for predicting bank underperformance: a comparison of neural networks, discriminant analysis, and professional human judgment. *Intelligent Systems in Accounting, Finance & Management*, *10*(3), 169-186.

[15] Kolari, J., Glennon, D., Shin, H., & Caputo, M. (2002). Predicting large US commercial bank failures. *Journal of Economics and Business*, *54*(4), 361-387.

[16] Lanine, G., & Vander Vennet, R. (2006). Failure prediction in the Russian bank sector with logit and trait recognition models. *Expert Systems with Applications*, *30*(3), 463-478.

[17] Marais, M. L., Patell, J. M., & Wolfson, M. A. (1984). The experimental design of classification models: An application of recursive partitioning and bootstrapping to commercial bank loan classifications. *Journal of accounting Research*, 87-114.

[18] Frydman, H., Altman, E. I., & Kao, D. L. (1985). Introducing recursive partitioning for financial classification: the case of financial distress. *The journal of finance*, *40*(1), 269-291.

[19] Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.

[20] Chen, W. H., & Shih, J. Y. (2006). A study of Taiwan's issuer credit rating systems using support vector machines. *Expert Systems with Applications*, *30*(3), 427-435.

[21] Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., & Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision support systems*, *37*(4), 543-558.

[22] Shin, K. S., Lee, T. S., & Kim, H. J. (2005). An application of support vector machines in bankruptcy prediction model. *Expert systems with applications*, *28*(1), 127-135.

[23] Kosmidou, K., & Zopounidis, C. (2008). Predicting US commercial bank failures via a multicriteria approach. *International Journal of Risk Assessment and Management*, *9*(1-2), 26-43.

[24] Gaganis, C., Pasiouras, F., & Zopounidis, C. (2006). A multicriteria decision framework for measuring banks' soundness around the world. *Journal of Multi-Criteria Decision Analysis*, *14*(1-3), 103-111.

[25] Niemira, M. P., & Saaty, T. L. (2004). An analytic network process model for financial-crisis forecasting. *International journal of forecasting*, *20*(4), 573-587.

[26] Ravi, V., & Pramodh, C. (2008). Threshold accepting trained principal component neural network and feature subset selection: Application to bankruptcy prediction in banks. *Applied Soft Computing*, *8*(4), 1539-1548.

[27] Pawlak, Z. (1982). Rough sets. *International journal of computer & information sciences*, *11*(5), 341-356.

[28] Greco, S., Matarazzo, B., & Slowinski, R. (1998, June). A new rough set approach to multicriteria and multiattribute classification. In *International conference on rough sets and current trends in computing* (pp. 60-67). Springer, Berlin, Heidelberg.

[29] Ahn, B. S., Cho, S. S., & Kim, C. Y. (2000). The integrated methodology of rough set theory and artificial neural network for business failure prediction. *Expert systems with applications*, *18*(2), 65-74.

[30] Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *6*(02), 107-116.

[31] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

[32] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2015, June). Gated feedback recurrent neural networks. In *International conference on machine learning* (pp. 2067-2075). PMLR.

[33] Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, *2*(3), 1-27.

[34] Fushiki, T. (2011). Estimation of prediction error by using K-fold cross-validation. *Statistics and Computing*, *21*(2), 137-146.