Journal of Theoretical and Applied Information Technology

15th December 2023. Vol.101. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

TIME SERIES CLASSIFICATION FOR FINANCIAL STATEMENT FRAUD DETECTION USING RECURRENT NEURAL NETWORKS BASED APPROACHES

ITO WASITO[1]*, FEBRYANTI SIMON[2], PRITA KARINA DIANDRA[2]

[1]Department of Informatics, Pradita University, Tangerang, Indonesia

[2]Department of Accounting, Pradita University, Tangerang, Indonesia

E-mail: [1]ito.wasito@pradita.ac.id, febryanti.simon@pradita.ac.id [2], prita.karina@pradita.ac.id[2]

[1*)] Corresponding Author

ABSTRACT

Financial statements fraud, considered as untruthful behavior for gaining financial benefits, has recently become a widespread issue in companies and organizations. With the advancement of artificial intelligence, deep-learning-based approaches can be used intelligently to detect fraudulent transactions by analyzing a large number of financial statements data. This paper proposes to formulate financial statements fraud detection into time series classification (TSC) problems using Recurrent Neural Networks (RNN) based approaches which include Simple RNN, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). The implementation of the proposed approaches will be demonstrated on financial statements data collected from Accounting and Auditing Enforcement Releases (AAERs), which are federal materials issued by US Security and Exchange Commission (SEC). The objective will be achieved through two steps. First, the Time Series Classification (TSC) and RNN based approaches will be reviewed. Then, the experimental settings of RNN based approaches comparisons on specified data sets will be introduced. As it is a TSC problem, therefore the evaluation of the RNN based model performances will be determined based on percentage of accuracy and loss-accuracy curves as visualization tool. This study has two contributions: first, this research can provide insight to corporate management and investors in detecting fraud that happens in the company, second contribution for the academic purpose, this research proposes alternative methods in detection of fraud. The results show that on various hidden unit number, GRU architecture model outperforms Simple RNN and LSTM. The best performance is achieved by GRU model at 14 hidden unit number which produces more than 99% in training accuracy and more than 91% in test accuracy. Overall, Simple RNN to be moderate model and LSTM is the worst model.

Keywords: *Time Series Classification, Fraud, Financial Statement, Recurrent Neural Networks, Deep Learning*

1. INTRODUCTION

Fraudulent financial reporting is defined as the intentional overstatement and or understatement of balances in the financial statements. Increasing fraudulent financial reporting among public companies in the past decade has focused public attention on the corporate financial reporting process [1]. Financial fraud can result in devastating consequences for the stability of a firm, as well as considerable losses to shareholders, industry, and even the whole market [4,10]. The global Association of Certified Fraud Examiners (ACFE) in 2020 showed that there was a loss of $3.6bn based on an examination of 2,504 cases from 125 countries [1].

In recent years, cases of fraud in the field of financial statements have been experienced by several large companies in various countries. For instances, fraud committed by Harris Scarf companies in Australia, Parmalat in Italy, Ahold in India, and Vivendi in France as well as in other well-known companies such as Enron, World Com, Tyo, and Lucent [1,2,3,4]. This phenomenon shows that the penetration on the fraud problem is so widespread throughout the world. Even though much effort has been made in relation to measuring the level of fraud in a company, obtaining accurate statistical data about fraud is not easy because the majority of fraud incidents are not detected. Indeed, even though indications of fraud have been found by a company, in general fraud victims prefer to cover up this information to avoid public attention. Furthermore, the increase of the occurrence of frauds in financial reports is generally closely related to the bankruptcy of a company, so it is necessary to encourage efforts to improve the quality of financial reports [1,4]

Several approaches to addressing the problem of fraud detection have been initiated in recent years. However, in general the methods used still use a manual approach and it is easy to understand that this type of approach is slow, expensive, inaccurate and less practical. Many attempts have been made to improve these deficiencies of the manual approach but the results have been unsatisfactory or inefficient.

With the advancement in the field of Artificial Intelligence, including several Machine Learning approaches as part of AI, it has been applied to detect fraudulent activity in the financial reporting sector. Deep Learning as the latest method in the field of AI has been widely used in fields including the financial sector with high accuracy and speed that exceeds several methods in other Machine Learning [2,3,4,11,20].

To handle the shortcomings of the traditional approach to detection of financial statement fraud and to utilize the advantages of Deep Learning approach, this paper will implement detection of fraud financial statement using Recurrent Neural Network (RNN) based approaches which includes: Simple RNN, Long-Short Term Memory and (LSTM) and gated Recurrent Unit (GRU). The advantages of RNN based approaches can be viewed as the detection model construction for financial statements time series data. Based on various time related data sets then the time series classification (TSC) model will be developed, finally an effective model detection of financial statements fraud will be established. The ultimate goal of this model is expected to reduce the losses of companies due to financial statement fraud and maintain the sustainable development of capital markets [14,15,16,17,18]. The objective of this paper is two-fold. First, the Time Series Classification will be reviewed and the Simple RNN, LSTM and GRU will be explored. Second, the performance of RNN, LSTM and GRU will be compared.

This study has two contributions: first, this research can provide insight to corporate management and investors in detecting fraud that happens in the company, second contribution for the academic purpose, this research proposes alternative methods in detection of fraud.

This paper will be organized as follows. First, Time Series Classification (TSC) problems will be reviewed. Next, the RNN based approaches for finance statements fraud detection will be explored. Then the experimental settings will be introduced. The results and discussions will show some findings from the experiments. The paper will be concluded by some recommendations for handling financial statements fraud detection.

2. TIME SERIES CLASSIFICATION

Time series classification (TSC) is defined as general time series analysis problem which requires capturing a functional dependence between the set of possible time series and the finites set of classes using a training set with known classes [[14,15,16,17,18]. There are several works which illustrate that deep neural networks are suitable for TSC problem and can outperform other algorithms which encourage further work in this direction. There exist deep learning models like recurrent neural networks (RNNs) that are designed specifically for processing sequential data, and thus could be applied for time series. Although some recent works use special kinds of RNNs as a component of resulting model for TSC problem, recent publications, which study the efficiency of RNNs as a self-sufficient approach for the problem, were not found.

From studies relying on traditional statistical methods, some literature uses data mining and machine learning techniques for the detection of financial statement fraud. In addition to better accuracy in prediction than traditional statistical methods, data mining and machine learning techniques produce more accurate classifications and predictions with machine learning on massive amounts of data [2,7,9,13]. Moreover, data mining and machine learning techniques do not need the presumptions required by traditional statistics and can effectively handle non-linear problems [3]. For example, methods such as Artificial Neural Network (ANN), Support Vector Machine (SVM), Decision Tree (DT), and Genetic Algorithm (GA) are being used to detect financial statement fraud [7,9,20]. In the era of artificial intelligence, deep learning techniques are used by studies for the detection of financial statement fraud [1]. Compared to machine learning, deep learning can better process massive volumes of data and have greater predictivity. While deep learning is based on neural networks, it has more layers than neural networks and can effectively capture features and handle complex issues. Compared to machine learning, it can more effectively detect financial statement fraud [3,4].

3. RECURRENT NEURAL NETWORKS AND THE EXTENSIONS

In this section, the mechanisms of the three types of Recurrent Neural Networks will be explored. The notation ⊕ and ⊗ refer to addition, Hadamard multiplication where in the equation written as ⊙. Sigmoid refer to the Sigmoid activation function.

3.1 Simple Recurrent Neural Networks

The simple recurrent neural network (RNN) is one of the most popular deep learning architectures and is often used for data analysis with a time related nature [14,15,16,17,18]. RNN is good at capturing the relations between each data point and can produce predictions on time series data with promising accuracy. Using RNN the important information from the past is preserved in the modeling process as the input variable to be computed with current data. The simple RNN architecture is illustrated in Figure 1. The computation result of the time *t* will be in a hidden state as the input variable. This input variable and the t + 1 data will be dispatched into the hidden layer together for computation. In general, the activation function used in the hidden layer of the recurrent neural network is a hyperbolic tangent function (tanh).

This computation expressed in (1) below.

$$y'_t = \tanh(w_{ht} y'_{t-1} + w_{xt} x_t + b_t) \qquad (1)$$

The RNN is commonly used for the capturing of temporal dynamic behavior sequences. The status is circulated and transmitted in its own network. It can accept a wider range of inputs in time-sequential structures such as text, audio and video. To distinguish from feedforward neural networks, RNN places a greater emphasis on network feedback process. In both scientific and real applications, RNN has demonstrated good results in solving problems such as voice to texts or translation [22].

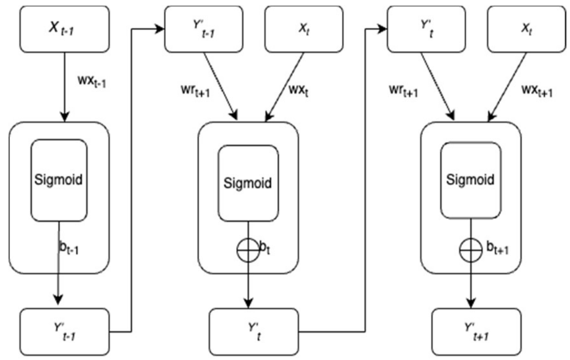

Figure 1. Architecture Of Simple Recurrent Neural Networks

In this paper, the RNN will be deployed for determining whether financial statements time series of companies can be classified as "fraud" or "non fraud".

3.2 Long Short Time Memory

The architecture of long short-term memory (LSTM) is an extension of the RNN based on the improvement introduced by Hochreiter and Schmidhuber [21] in 1997 to overcome the weak performance of long-term memory situation in RNN. LSTM has extension of three steps (i.e., forget, update, and input) in the neurons, to significantly improve the long-term memory performance. On the basis of RNN, LSTM replaces the neurons in the hidden layer of RNN with a gate control mechanism and memory cells, so as to overcome the problem of gradient disappearance and poor long-term memory effects of RNN. The gate control mechanism consists of Forget Gate (FG), Input Gate (IG), and Output Gate (OG) as shown in Figure 2. Forget Gate determines the portion to be forgotten from the input of the previous period to the memory status of the current period, as expressed in (2). Input Gate determines the input variable during the current period and the output result from the previous data to be updated to the memory status, as expressed in (3).

The update is expressed in (4). Output Gate determines the output information from the memory status, as expressed in (5). The final calculation of the input value is expressed in (6).

$$f_t = sigmoid[w_f\, y'_{t-1} + x_t] + b_f) \qquad (2)$$

$$i_t = sigmoid\left[w_f\left(y'_{t-1} + x_t\right)\right] + b_i \quad (3)$$

$$L_t = L_{t-1} * f_t + \tanh[w_L\,(y'_{t-1} + x_t) + b_L) * i_t \quad (4)$$

$$o_t = sigmoid\ [w_o(y'_t + x_t)\} + b_0) \quad (5)$$

$$y'_t = \tanh(L_t) * o_t \quad (6)$$

Similar to the RNN, the LSTM also consists of repeated and interconnected units. However, the design of internal units is more flexible as they interact with each other. RNN and LSTM are powerful in the reception of sequences of either single or multiple vectors for input data or output data. When the vector input of a fixed size only allows the output of single vectors, the network can process classification problems. In the experiments, the output of single vectors either “fraud” or “non fraud” financial statements.

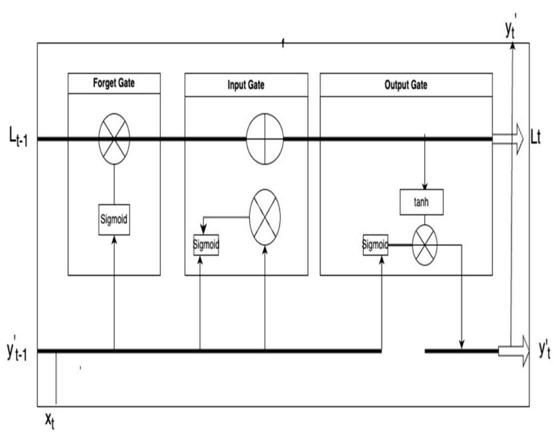

Figure 2. The Architecture Of Long Short Time Memory

3.3 Gated Recurrent Unit

The Gated Recurrent Unit is the extension of RNN architecture using Encoder Decoder which consists of two RNN [22]. GRU (Gated Recurrent Unit) aims to solve the vanishing gradient problem which comes with a standard recurrent neural network. GRU can also be considered as a variation on the LSTM because both are designed similarly and, in some cases, produce equally excellent results [5,6]. The general GRU architecture shown in Figure 3.

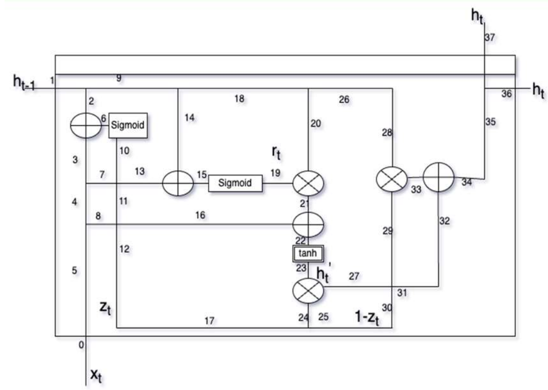

Figure 3. Architecture Of Gated Recurrent Unit

In order to solve the vanishing gradient problem of a standard RNN, GRU uses, so-called, *update gate* and *reset gate* which will be explored more details later. Basically, these are two vectors which decide what information should be passed to the output. The special thing about them is that they can be trained to keep information from long ago, without washing it through time or remove information which is irrelevant to the prediction. This computation model is in favor for time series classification especially in the context of financial statements fraud detection.

3.3.1 Update Gate

In this stage, the process at GRU start by computing the update gate z_t for time step *t* using the following formula:

$$z_t = \sigma(w_z x_t + u_z h_{t-1}) \quad (7)$$

When x_t is attached into the network unit, it is multiplied by its own weight w_z. The same process for h_{t-1}which holds the information for the previous $t-1$ units and is multiplied by its own weight u_z. Both results are added together and a sigmoid activation function is applied to squash the result between 0 and 1. The gate is represented by following index of line number: 1-2-⊕-0-5-4-3-⊕-6-*Sigmoid*-10-11-12-17-24-⊗-23-22.

The update gate leads the model to determine how much of the past information (from previous time steps) needs to be passed along to the future. That is really powerful because the model can decide to copy all the information from the past and eliminate the risk of vanishing gradient problem. The usage of the update gate will be explored later on.

3.3.2 Reset Gate

Essentially, this gate is used from the model to decide how much of the past information to forget. To calculate it, we use:

$$r_1 = \sigma(w_r x_t + u_r h_{t-1)} \qquad (8)$$

This formula is the same as the one for the update gate. The difference comes in the weights and the gate's usage, which will see in a bit. As before, we plug in h_{t-1} and x_t, multiply them with their corresponding weights, sum the results and apply the sigmoid function. This gate represented by index of line number as follows: 1-9-14-⊕-5-4-7-13-⊕-15-*Sigmoid* -19.

3.3.3 Current Memory Content

Let's see how exactly the gates will affect the final output. First, the process starts with the usage of the *reset gate*, then a new memory content which will use the reset gate to store the relevant information from the past. It is calculated as follows:

$$h_t' = \tanh(w_{x_t} + r_t \odot uh_t) \qquad (8)$$

The step can be traced using the following index of line number: 1-9-18-20-⊗-19-sigmoid and ⊗-21-⊕-22-tanh-23-and 36-37.

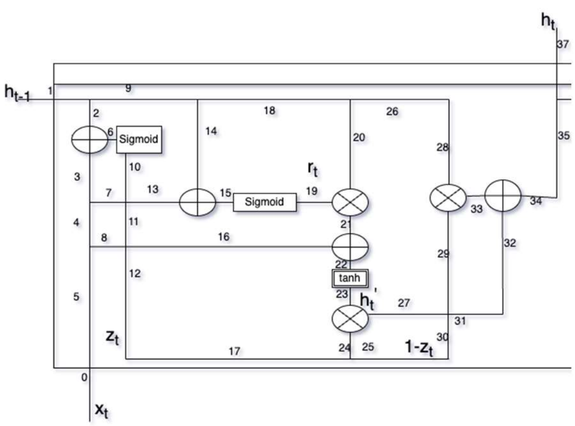

Figure 3. Architecture Of Gated Recurrent Unit

3.3.4 Final Memory at Current Time Step

As the last step, the network needs to calculate $h_t$vector which holds information for the current unit and passes it down to the network. In order to do that the update gate is needed. It determines what to collect from the current memory content h_t' and what from the previous steps h_{t-1}. This is shown in the following equation:

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot h_t' \qquad (9)$$

This operation can be traced through the following line number index: 0-1-9-18-26-28-⊗-33-⊕-34-35-37 and 10-11-12-17-24-⊗-23 and 25-30-29-⊗.

4. EXPERIMENTAL SETTINGS

4.1 Time Series of Financial Statements Data

The financial statements data collected from Accounting and Auditing Enforcement Releases (AAERs), which are federal materials issued by US Security and Exchange Commission (SEC) at https://www.sec.gov/ for alleging accounting and/or auditing misconduct in U.S.-listed companies. The twenty-three companies are selected such that they have full yearly report K-10 form between year 2012 to 2020. Among twenty-three companies there are seven companies confirmed to have fraud issues.

For the analysis purposes, the financial statements will be arranged into time series data structure. The next step, the variables will be used to selected. Those variables which are available for all selected companies including: *Cash and Cash Equivalents at Carrying Value, Assets Current, Liabilities Current, Liabilities And Stockholders Equity, Net Cash Provided By Used In Operating Activities*. Thus, there are 3-D data sets.

4.2 Evaluation of Performances

The RNN model and its extensions, LSTM and GRU will be compared for financial statements fraud detection using the specified data sets as described above. The comparisons will utilize three number of hidden layer units which include: 6, 10 and 14.

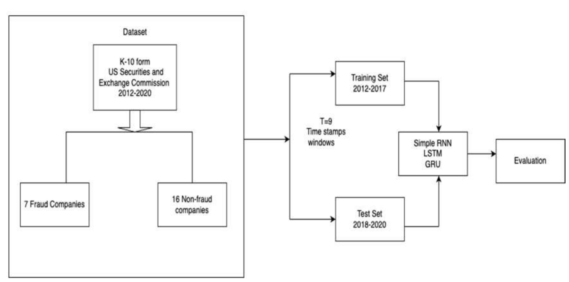

Figure 4. The Model Evaluation Using T Stamp Windows

Journal of Theoretical and Applied Information Technology
15th December 2023. Vol.101. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The next stage, data will be divided into two sets: training set for generating predictive model and test set for evaluating the performances of RNN, LSTM and GRU in detecting of financial statements data fraud based on the accuracy of the model prediction on training sets and test sets.

The loss curves and validation curves will used to visualize the goodness of the prediction model of simple RNN, LSTM and GRU for detecting of financial statements data fraud.

5. RESULTS AND DISCUSSION

5.1 Accuracy of Fraud Detection

As shown in table 1, overall, for all three RNN based architectures as the hidden unit increased, the train accuracy will be increased except for LSTM using 14 hidden units. Similar results are shown for test accuracy with the exception for LSTM which provides the similar results for all hidden unit numbers. Furthermore, Simple RNN surprisingly has better performance compare to the LSTM but still has lower percentage train and test accuracy to the GRU.

Table 1. Experimental Comparisons Between Simple RNN, LSTM and GRU

	Simple RNN			LSTM			GRU		
#Hidden Unit	6	10	14	6	10	14	6	10	14
Train Accuracy(%)	79.23	93.85	96.92	73.85	77.69	73.85	80.00	81.54	99.23
Test Accuracy (%)	65.22	88.41	91.30	69.57	69.57	69.57	78.26	81.16	92.76

Overall, GRU seems outperforms to the Simple RNN and LSTM. Especially at 14 hidden unit case, GRU is clearly better than Simple RNN. In contrast, the results show that LSTM consistently to be the worst. It may indicate that LSTM is not suitable for short time series classification. The nature of LSTM as explored in section 3.2 will be fit for longer time series classification problems especially for financial statements data fraud detection. From table 1, the results show that GRU is a recommended model for financial statements fraud detection.

5.2 Loss and Accuracy Validation

To see the goodness of the three RNN based models, the loss and accuracy are validated as shown in Figure 5 to Figure 13.

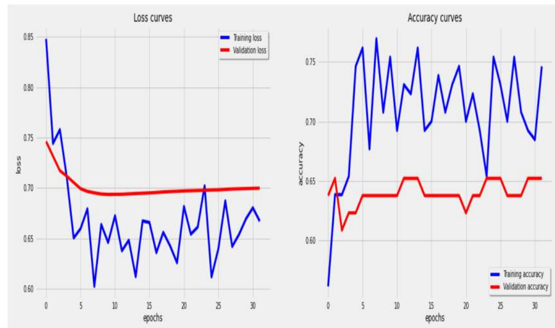

Figure 5. Simple RNN With 6 Hidden Units

Overall, the results from table 1 is confirmed by loss and accuracy curves. For Simple RNN as shown in Figure 5- Figure 7, the results show that Simple RNN has better performance at 10 and 14 hidden units. In contrast, at 6 hidden unit number, Simple RNN has the worst results.

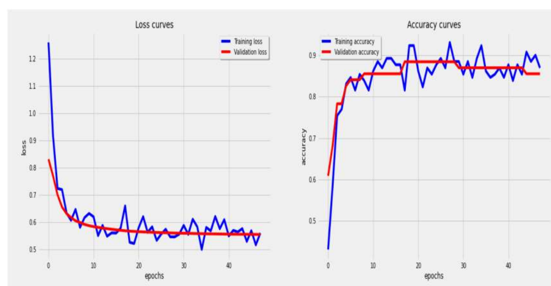

Figure 6. Simple RNN With 10 Hidden Units

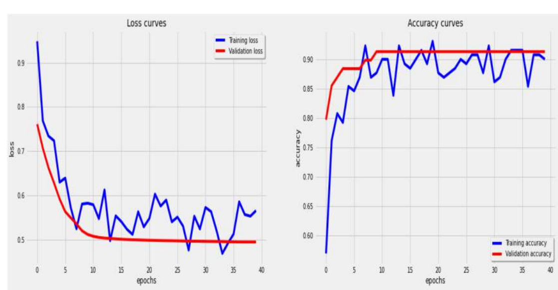

Figure 7. Simple RNN With 14 Hidden Units

Figure8-Figure10 show that LSTM consistently has poor performance for detecting of fraud of financial statements. Overall, LSTM is

unstable performances when the hidden unit increased.

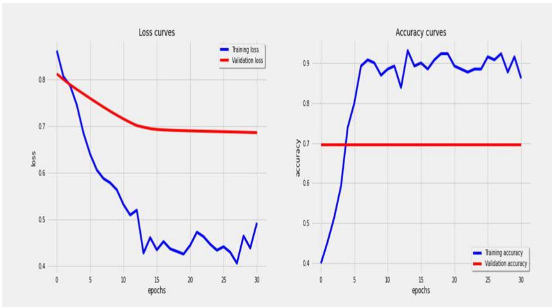

Figure 8. LSTM With 6 Hidden Units

Inline the results from Table 1, LSTM produces similar performances in the validation accuracy as shown in Figure 8-10.

As expected, the GRU as shown in Figure 11 – Figure 13, produces promising results. When GRU uses 6 hidden units, performance of GRU model is not so good.

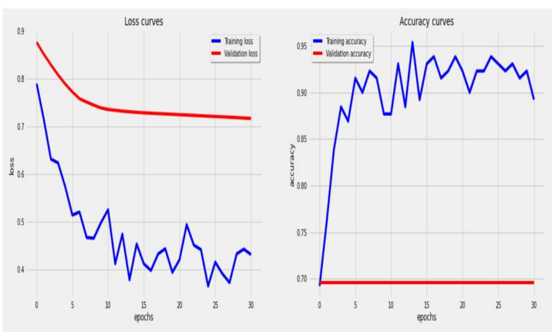

Figure 9. LSTM With 10 Hidden Units

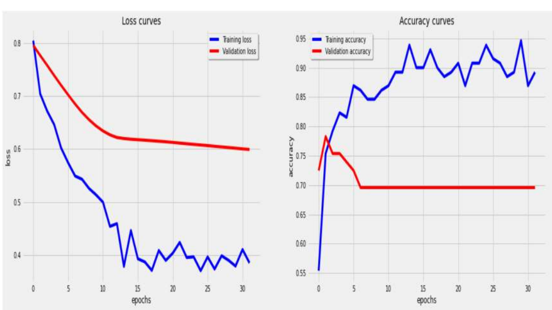

Figure 10. LSTM With 14 Hidden Units

However, when the number of hidden units increased, the performance of GRU is also improved. Especially at 14 hidden units, the GRU reach the best performance and outperforms the other RNN based approaches as shown in Figure 13.

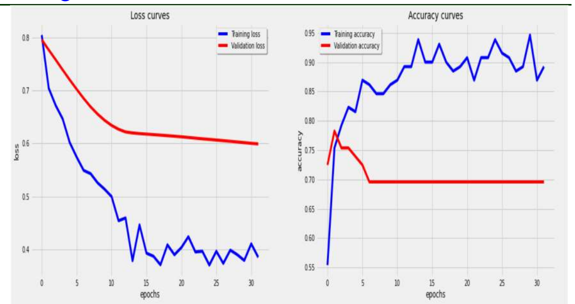

Figure 11. GRU With 6 Hidden Units

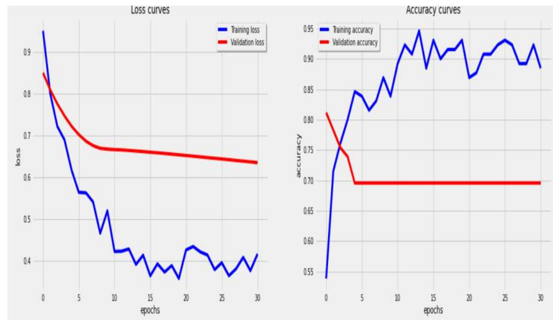

Figure 13 GRU With 10 Hidden Units

However, when the number of hidden units increased, the performance of GRU is also improved. Especially at 14 hidden units, the GRU reach the best performance and outperforms the other RNN based approaches as shown in Figure 14.

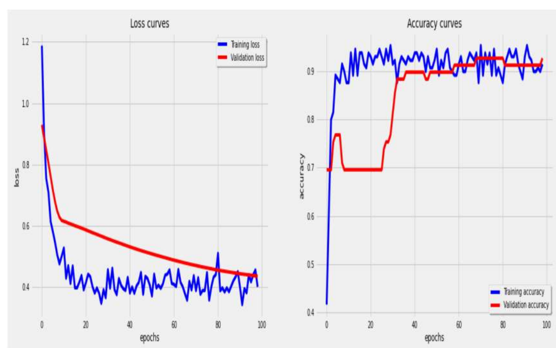

Figure 13. GRU With 14 Hidden Units

In general, the GRU has promising performance and outperforms other methods as proposed in [1,2].

6. CONCLUSIONS

The report has two contributions; first, research can provide insight to corporate management and investors in detecting fraud that happens in the company, second contribution for the academic purpose, this research propose alternative methods in detection of fraud.

In this report, the problem of financial statements data fraud data detection is formulated as Time Series Classification problems. To handle the problems, RNN based approaches had been implemented using real data sets from US Security and Exchange Commission (SEC).

First, the three main RNN based approaches Simple RNN, LSTM and GRU were reviewed. Then, some experimental settings are designed to compare performances of Simple RNN, LSTM and GRU using three types of hidden unit which include: 6, 10 and 14 in the experiments. The evaluation of performances is evaluated using training and test accuracy as well as loss and accuracy curves as visualization tools.

Overall, the results show that when the hidden unit number increased the performances of the model improved. The GRU model, except for 10 hidden unit number, consistently outperforms Simple RNN and LSTM. In contrast, the results show that LSTM model to be the worst for all types of hidden unit number.

From above results, the GRU model or Simple RNN model is a recommended approach for short time series data, the LSTM is not suitable for this type of time series data.

However, the theoretical research of the proposed methods needed to be explored in the future.

7. ACKNOWLEDGEMENT

We are grateful to the Pradita University for providing financial support using Internal Cooperation of Study Program Research Grant No. 003/LPPMPT/SPK/Pradita/VI/2023.

REFERENCES

[1] Jan, Chyan-Long. "Detection of Financial Statement Fraud Using Deep Learning for Sustainable Development of Capital Markets under Information Asymmetry" *Sustainability* 13, no. 17: 9879, 2021.

[2] Chen, S. "Detection of fraudulent financial statements using the hybrid data mining approach," *SpringerPlus* **5**, 89, 2016.

[3] Ali, Abdulalem et al., "Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review", *Applied Sciences* 12, 2022.

[4] Dong, Wei et. al., "The Detection of Fraudulent Financial Statements: An Integrated Language Model", in PACIS 2014 Proceedings. 383, 2014.

[5] R. Dey and F. M. Salem, "Gate-variants of Gated Recurrent Unit (GRU) neural networks," *in IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, Boston, MA, USA, 2017, pp. 1597-1600, 2017.

[6] Guizhu Shen, Qingping Tan, Haoyu Zhang, Ping Zeng, Jianjun Xu,"Deep Learning with Gated Recurrent Unit Networks for Financial Sequence Predictions", Procedia Computer Science, Volume 131, pp 895-903, 2018.

[7] Petr Hajek, Roberto Henriques, "Mining corporate annual reports for intelligent detection of financial statement fraud – A comparative study of machine learning methods", Knowledge-Based Systems,Volume 128, pp 139-152. 2017.

[8] Hayashi, A., Mizuhara, Y., Suematsu, N. Embedding Time Series Data for Classification. In: Perner, P., Imiya, A. (eds) Machine Learning and Data Mining in Pattern Recognition. MLDM 2005. Lecture Notes in Computer Science (), vol 3587. Springer, Berlin, Heidelberg. 2005.

[9] J. West and M. Bhattacharya, "Mining financial statement fraud: An analysis of some experimental issues," in *2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA)*, Auckland, New Zealand, pp. 461-466, 2015.

[10] Zaki, Mohamed and Zaki, Mohamed and Theodoulidis, Babis, "An Ontology for Monitoring and Surveillance Systems in Financial Markets" (April 17, 2013). Available at SSRN: https://ssrn.com/abstract=2266679 or http://dx.doi.org/10.2139/ssrn.2266679

[11] Lu, Yi-Wei, Chia-Yu Hsu, and Kuang-Chieh Huang., "An Autoencoder Gated Recurrent Unit for Remaining Useful Life Prediction", *Processes* 8, no. 9, 2020.

[12] Lipton, Zachary Chase. "A Critical Review of Recurrent Neural Networks for Sequence Learning." *ArXiv* abs/1506.00019, 2015.

[13] Ajit Kr. Singh Yadav and Marpe Sora, "Fraud Detection in Financial Statements using Text Mining Methods: A Review" in 2021 *IOP Conf. Ser.: Mater. Sci. Eng.*, 2021.

[14] Michael Hüsken, Peter Stagge, "Recurrent neural networks for time series classification",

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195
Neurocomputing, Volume 50, pp. 223-235, 2003.

[15] Santos, Tiago and Roman Kern. "A Literature Survey of Early Time Series Classification and Deep Learning", in *SAMI@iKNOW*, 2016.

[16] Ismail Fawaz, H., Forestier, G., Weber, J. *et al.* Deep learning for time series classification: a review. *Data Min Knowl Disc* **33**, pp. 917–963, 2019.

[17] Smirnov, Denis and Mephu Nguifo, Engelbert. *"Time Series Classification with Recurrent Neural Networks", unpublished,* 2018.

[18] *Foumani, Navid, et al. "Deep Learning for Time Series Classification and Extrinsic Regression: A Current Survey", Arxiv.org/abs/2302.02515, 2015.*

[19] Fletcher H. Glancy, Surya B. Yadav,*,"A computational model for financial reporting fraud detection"*, Decision Support Systems, Volume 50, Issue 3,pp 595-601, 2011.

[20] Mohammadi, M. A., Shohreh Yazdani and Mohammad Hamed Khanmohammadi. "Presenting a Model for Financial Reporting Fraud Detection using Genetic Algorithm." Advances in Mathematical Finance & Applications, vol. 6 no.2 , pp.377-392, 2021.

[21] Hochreiter, S.; Schmidhuber, J. "Long short-term memory", Neural. Comput.. 1997, 9, 1735–1780

[22] Cho, K et al. "Learning Phrase Representations using RNN Encoder–Decoderfor Statistical MachineTranslation", *Arxiv.org/abs/1407.1078v3, 2014*