# Deep learning for Predicting Milk Volume in Taiwan

Kao Sheng Che, Wu Te Wei, Ko Yen Tsun

*Dept. Communication Engineering, National Tsing Hua University of Taiwan*

This work was supported by the Final project of Prof. Chia Wen Lin Deep learning course for supporting this research result via the Aldea Platform of ITRI

***Abstract*— Deep learning (DL) algorithms are considered as a methodology of choice for analysis over the past few years. According to the project. First, we use some methods in our pre-processing, by Random Forest Tree and PCA to find our training feature. Secondly, we present the details of the feature and conduct them into new features. Then we focus on label-encoding and one hot encoding to revise data before training. Thirdly, we use different model in our training process, these deep learning-based methods were classified into three model; SVM methods, LSTM methods and GRU Method to compare which has the best result This study will contribute to different method understanding of deep learning for change method and provide a basis for further research.**

***Keywords*— Deep learning, PCA, Random Forest Tree, One hot/Label encoding, SVM, LSTM ,GRU**

## I. INTRODUCTION

Deep learning (DL) has seen an increasing trend and a great interest over the past decade due to its powerful ability to represent learning. Deep learning allows models that are built, based on multiple processing layers, to learn representations of data samples with several levels of abstraction. In the Aldea platform, we can get the four files report.csv, birth file, breed file, spec file. Report file uses these features such as Parity、Days of lactation、Number of mating、Dairy Farm ID、Data of Month。Birth file uses these features such as Female Cow of Weight、Times。Spec file uses a feature named Status。We don't use breed files according to our pre-processing. It can't help add the training accuracy. After combining these files by Cow ID and some data processing, the final training dataset will have only 10 features. Then, try to use different model and also test the parameter in these model to find the best accuracy.

## II. RELATED WORK

A. *PCA*(Negash, 2021)*: Principal component analysis (PCA) is the process of computing the principal components and using them to perform a change of basis on the data, sometimes using only the first few principal components and ignoring the rest.*

B. *Random Forest Tree(*(Okada et al., 2021)*: Random forests are method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean/average prediction (regression) of the individual trees.*

C. *One hot/ Label Encoding*(Zhou et al., 2021)*:*

D. *SVM: A Support Vector Machine ( is a supervised machine learning algorithm that can be employed for both classification and regression purposes*

E. *LSTM: An LSTM has a similar control flow as a recurrent neural network. It processes data passing on information as it propagates forward*

F. *GRU: The GRU is the newer generation of Recurrent Neural networks and is pretty similar to an LSTM GRU's got rid of the cell state and used the hidden state to transfer information*

## III. METHOD

Our Method will divide into five step. First we will use two methods (PCA and Random Forest Tree) to discuss which feature we can use. Then we will do some process on those data. Finally, use different model to train the accuracy

### A. *Framework*

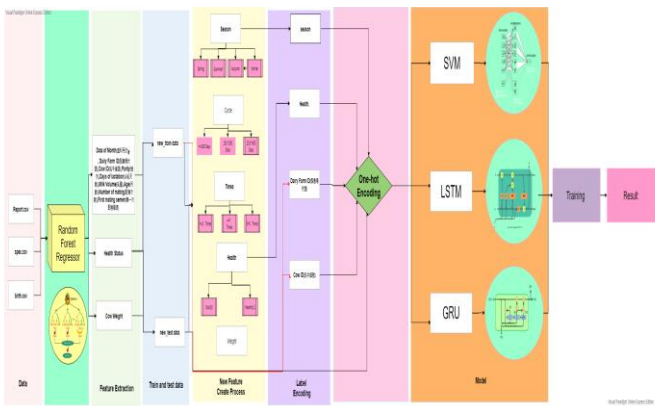


*Fig. 1 Model flowchart*

- *Feature Extraction*

According to our Random Forest Regressor Result, We choose the Data of Month, Dairy Farm ID, Cow ID ,Parity, Days of lactation, Milk Volume ,Age ,Number of mating ,First mating semen from the report file Then use the Female Cow of Weight from birth file Besides, we also distinguish the cow health status from spec file

- *New feature process*

Create new season','cycle','times','health','weight column in our train dataset

- *Season-Use the month from report to divide into 4 season Spring( 3~5), Summer( 6 ~8), Autumn( 9 ~11), Winter( 12 ~2)*
- *Cycle-Use the Days of lactation to to divide them into three cycle Cycle1(<= 100 Days), Cycle 2 (101~200) Cycle 3(201~ 305)*
- *Times Use the Parity to divide them into three times Time1(<= 2) ,Time 2(>=7), Time3(3~6)*
- *Health Use the Health Status to divide them into two cases Sick( 0), Health( 1)*

- *Label encoding*

We will use label encoding to let different categorical to the number, so we use this way on Dairy Farm ID, Cow ID, Season, Health

- *One hot encoding*

After we use the label encoding, we will use the one hot encoding to produces a vector with length equal to the number of categories in the dataset

- *Model*(Garg & Biswas, 2021)

We choose the three model in our training, they are SVM, LSTM, GRU. All of them Optimizer use ADAM. Learning rate initial is 0 001. Batch size is 64 .Max epochs is 30

### B. *Feature Extraction*

- *PCA*

A property of PCA is that you can choose the number of dimensions or principal component in the transformed result.

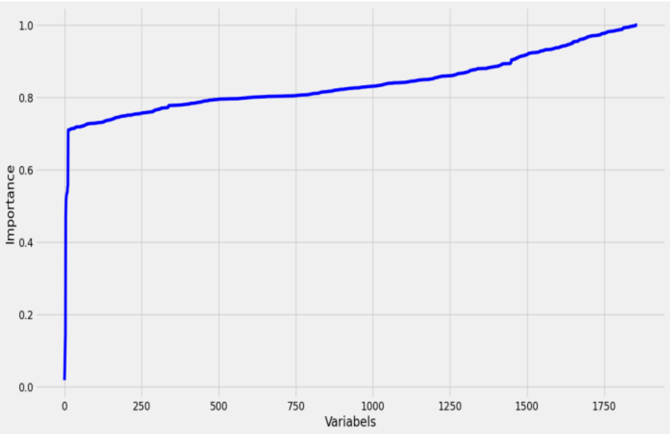


FIG. 2 PCA RESULT

The figure represents the importance of each features. The summary of each features' importance is 1. Importance's describe how important features are for the machine learning model. After the computation from the Random Forest structure, we can select the lower dimension of the training dataset.

- *Random Forest*

At each node (this is at each question), the three divide the dataset into 2 buckets, each of them hosting observations that are more similar among themselves and different from the ones in the other bucket. Therefore, the importance of each feature is derived from how "pure" each of the buckets is

| Variables | Importance |
|---|---|
| Days of lactation | 0.32 |
| Dairy Farm ID | 0.17 |
| Cow ID | 0.05 |
| Times | 0.04 |
| Age | 0.04 |
| Date of year | 0.03 |
| Date of month | 0.02 |
| Number of mating | 0.01 |
| First mating semen | 0.01 |

TABLE1 RANDOM FOREST RESULT

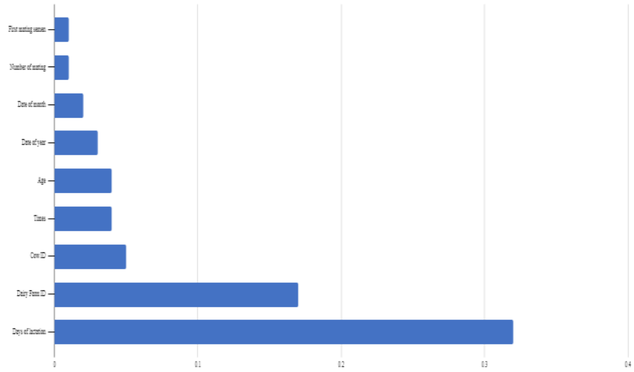

FIG. 3 RANDOM FOREST DISTRIBUTION

### C. *Model*

#### 1) *SVM*((Zeng et al., 2021)

The model we use as the beginning was SVM.The parameters of SVM are set as followings: training epoch is 25, batch size is 64,activation function is relu. The best result of rmse loss is 5.6552.However,the result doesn't increase after we adjust the parameters. Therefore, we change our model from SVM to LSTM ,in order to find out the better results

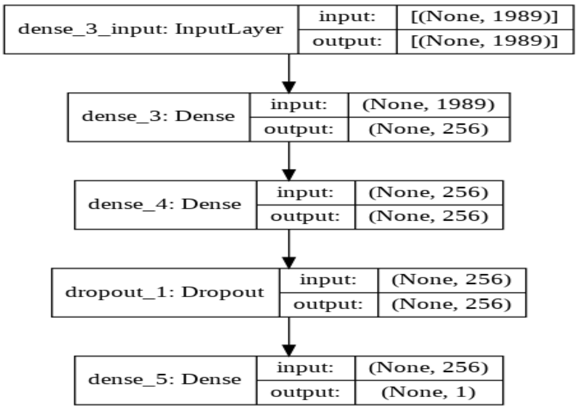


FIG. 4 SVM ARCHITECTURE

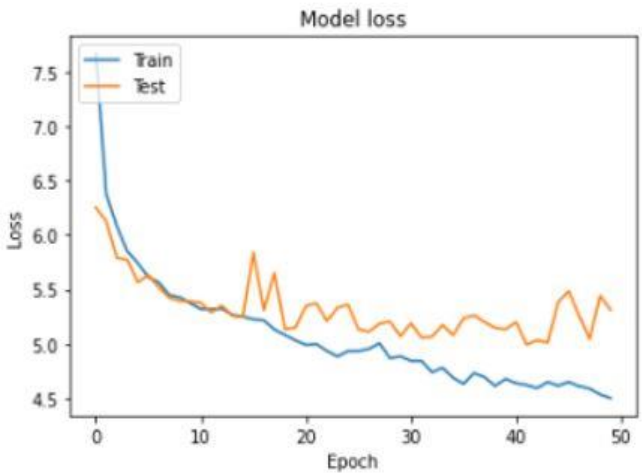


Fig5. SVM Model Loss(rmse)

2) *LSTM*(Weng et al., 2021)

After we tried the SVM model, we then tried the LSTM model. We experimented with "adding orthogonal initialization in different layers", "increasing Epochs", "increasing the number LSTM' node"

Our LSTM model architecture:1 LSTM Layer with 300 nodes,1 Dense Layer with 200 nodes ,1 Dropout Layer (0.2), then 1 dense Layer with 1 node。 With 20 epochs.

Due to gradient decent and gradient vanish in training LSTM, add orthogonal initialization in different layers.There are three situations in our result shown below. The result of testing sets are well-fitting by models. The loss of LSTM is more lower than SVM and GRU. Finally, we choose LSTM as our model to train.

-Without orthogonal initialization

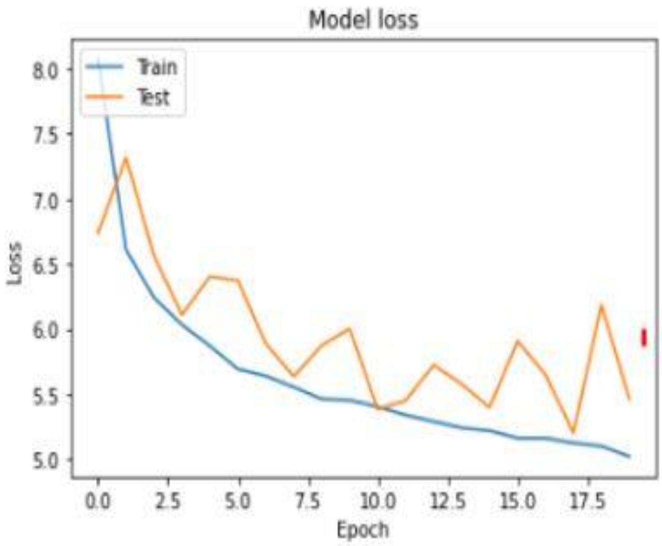


Fig6 . Without orthogonal initialization

- Adding to all layers

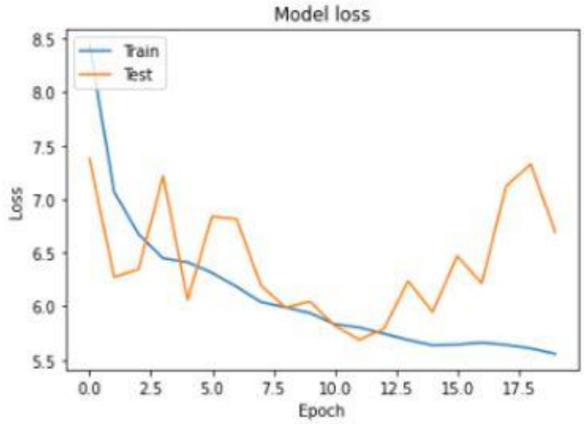


Fig7. Adding to all layers

-Adding to the LSTM layer only.

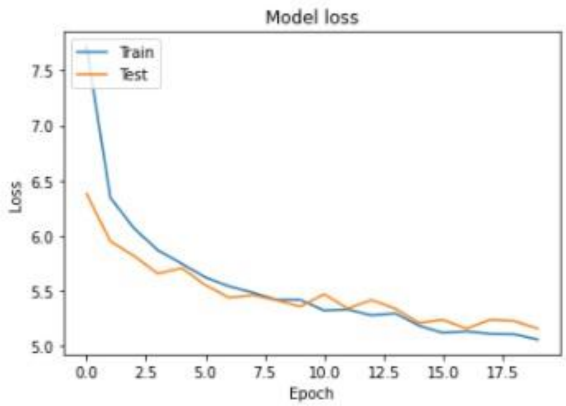


Fig8. Adding to the LSTM layer only.

Because we also try to test the epoch from 20 to 50, the result of epoch=50 would overfitting, we finally choose Epoch=20 on our training Then change the node with the LSTM This way of conditions have similar performance.

3) GRU((Xuan et al., 2021)

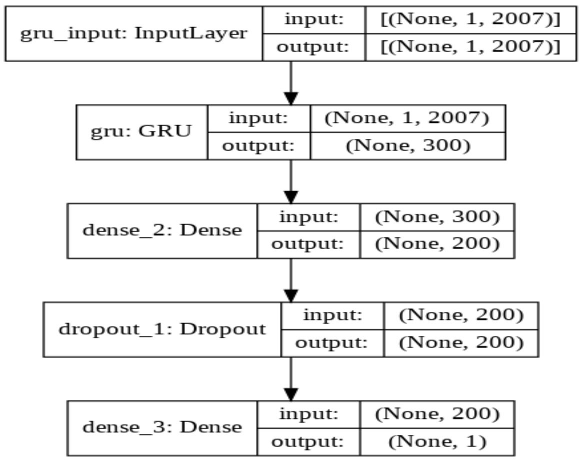


Fig9. GRU Architecture

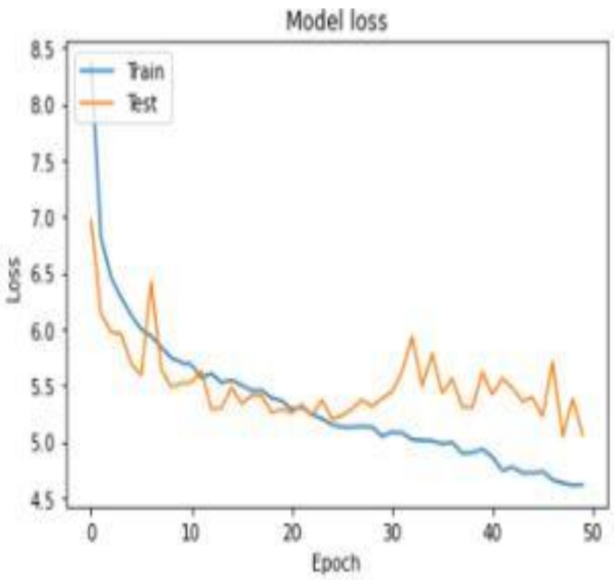


Fig10. GRU Model Loss(rmse)

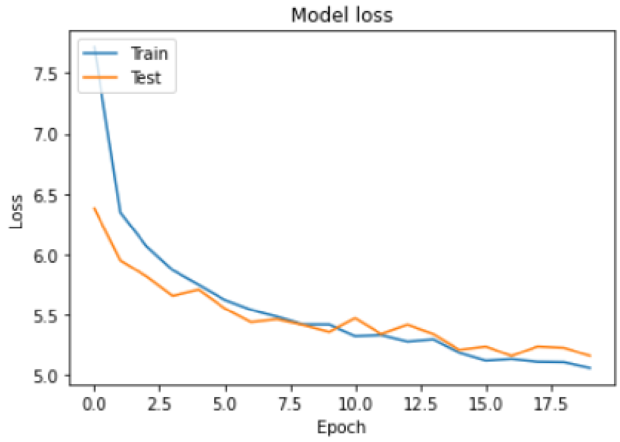


Fig12 .LSTM Model RMSE Loss(best)

| Method | RMSE Loss |
|---|---|
| SVM | 5.665 |
| LSTM | 5.539 |
| GRU | 5.977 |

Table 2.Comparison of the proposed method with other methods

Despite the models of GRU being much simpler and faster training than LSTM, we discover the fitting results of GRU are worse than LSTM with epochs of 50 by the experiments.

### *D. Result*

According to our final result in the different model, we choose the best model-LSTM,

The LSTM Layer input is 300, Dense Layer is 200. other parameter Optimizer use ADAM. Learning rate initial is 0 001. Batch size is 64. Max epochs is 25. To sum up, our experiments reduce RMSE to 5.539.

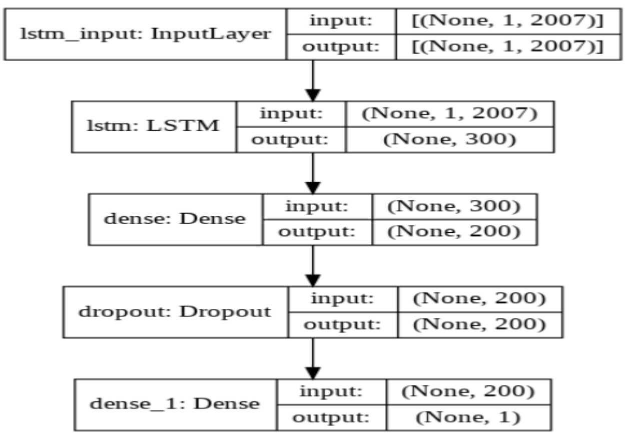


Fig11 .LSTM Architecture(best)

## IV. Conclusions

We use three methods to train our model, The first method SVM is the conventional method, it's also our first way to build the complete model to see our pre-process of data. Then, we also adopt the TA suggestion to use the LSTM train model. It's better than our previous SVM model. Besides, we also use the new way-GRU which is similar to the LSTM. Although the result we experimented with didn't work better than the result, we found its process time is shorter than LSTM. To sum up, our experiments reduce RMSE to 5.539.

In the future, maybe we still try to adopt Gradient Boost and Ensemble learning in our model to experiment this way can increase our model accuracy.

## Acknowledgment

The final project would like to thank the Aldea and Deep Learning course of TA and teacher for having supported this final project via the group competition program.

## References

[1] Garg, G., & Biswas, M. (2021). Improved Neural Network-Based Plant Diseases Identification. In *Lecture Notes in Electrical Engineering* (Vol. 668). https://doi.org/10.1007/978-981-15-5341-7_6

[2] Negash, F. (2021). Application of principal component analysis for predicting body weight of Ethiopian indigenous chicken populations. *Tropical Animal Health and Production*, *53*(1). https://doi.org/10.1007/s11250-020-02526-w

[3] Okada, Y., Matsuyama, T., Morita, S., Ehara, N., Miyamae, N., Jo, T., Sumida, Y., Okada, N., Watanabe, M., Nozawa, M., Iiduka, R., & Ohtsuru, S. (2021). Machine learning-based prediction models

for accidental hypothermia patients. *Journal of Intensive Care*, *9*(1). https://doi.org/10.1186/s40560-021-00525-z

[4] Weng, Z., Li, W., & Jin, Z. (2021). Human activity prediction using saliency-aware motion enhancement and weighted LSTM network. *Eurasip Journal on Image and Video Processing*, *2021*(1). https://doi.org/10.1186/s13640-020-00544-0

[5] Xuan, W., Shouxiang, W., Qianyu, Z., Shaomin, W., & Liwei, F. (2021). A multi-energy load prediction model based on deep multi-task learning and ensemble approach for regional integrated energy systems. *International Journal of Electrical Power and Energy Systems*, *126*. https://doi.org/10.1016/j.ijepes.2020.106583

[6] Zeng, J., Zou, B., Qin, Y., Chen, Q., Xu, J., Yin, L., & Jiang, H. (2021). Generalization ability of online pairwise support vector machine. *Journal of Mathematical Analysis and Applications*, *497*(2). https://doi.org/10.1016/j.jmaa.2020.124914

[7] Zhou, K., Li, S., Zhou, X., Hu, Y., Zhang, C., & Liu, J. (2021). Data-driven prediction and analysis method for nanoparticle transport behavior in porous media. *Measurement: Journal of the International Measurement Confederation*, *172*. https://doi.org/10.1016/j.measurement.2020.108869