A Comparative Study on Facial Recognition between Convolutional Neural Network and Recurrent Neural Network Speeds

Jasper R. Sunga and Val Randolf M. Madrid

***Abstract*— This study compares two types of neural networks, specifically, Convolutional Neural Network (CNN) and Reccurent Neural Network (RNN), in its simplest form to record the attendance of students using face recognition given that there are no pre-existing models and dataset that can be used during the first day of classes. The study shows that the CNN model used is better than the RNN model in image classification. The study also shows that the RNN model used is not feasible for this type of scenario. To achieve a consistent accuracy of at least 90%, the CNN model needs to be trained for a total of six days worth of dataset.**

***Index Terms*— face recognition, cnn, rnn, attendance, image classification**

I. Introduction

A. Background of the Study

Classes within schools or universities have attendances of students monitored. It is important that the attendance is monitored as it together with peer effect has a direct correlation to academic performance [1]. Furthermore, the performance of students in their academics shows that low performing students are affected by absenteeism compared to high performing students [2]. The traditional methods of monitoring the attendance are roll call, and manual checking of designated seats. These methods of checking the attendance consumes a significant amount of time which results to lesser lecture times.

Alternative ways in checking the attendance that aims to solve the issue of the traditional methods include the use of fingerprint-based biometrics systems and radio frequency identification system or RFID. A fingerprint-based biometrics system is a commercially available system. It reads the employee's fingerprint and records the fingerprint for the identification of the said employee. Improvements on the system, specifically using the optimized 5/3 DWT architecture and CORDIC Based FFT which are used for scanning and comparing images, were found to make the system more accurate [3]. RFID systems use tags or contact pads that has the information of an employee attached. The RFID-based attendance system made by Al-Naima and Ameen for Al-Nahrain University improved the traditional recording of attendance, roll call, as it takes into consideration the time, reliability, efficiency and ease of control to make the system [4].

Presented to the Faculty of the Institute of Computer Science, University of the Philippines Los Baños in partial fulfillment of the requirements for the Degree of Bachelor of Science in Computer Science

The fingerprint biometrics system and the radio frequency identification (RFID) systems are already established but the cost of the mentioned systems should still be considered. In 2017, the cheapest USB fingerprint scanner costs $50.00 or around PhP 2500.00 [5]. Currently, the minimum cost of a fingerprint biometrics system's hardware is still at around PhP 1500.00 on online stores. Depending on its capabilities and speed, RFID systems can be expensive from equipment cost to licensing cost. Passive RFID systems will cost you at least $1000.00 or PhP50000.00 for the reader and $.10 or PhP5.00 for the tag while active RFID systems will cost you at most 10 times less for the system compared to the passive system but will cost you at least $5.00 or around PhP250.00 for the tags [6].

With the advancement of image processing in computer science, face recognition can now be achieved in any computers and smartphones. Edge detection and feature extraction methods had become a way to achieve face recognition. The different methods can achieve accurate results with any given environments surrounding the face and angles of the camera [7]. With the advancement of face recognition and the right combinations of filters, edge detection algorithms, and features extraction methods, a system that will monitor the attendance of students in a class can be done.

B. Statement of the Problem

The traditional method of checking attendance consumes a significant amount of time given that the average duration of a class is one hour and thirty minutes. With the advancement of technology, face recognition can be used to check the attendance of a class. However, initial data of classes is usually required to be inputted first before face recognition methods can be used. Thus, possible data privacy issues may arise for some students.

C. Significance of the Study

The study seeks to solve the problem in finding the most efficient neural network architecture in using face recognition for attendance checking in classes. It also seeks to solve the problem of having a global database of students' faces before a working face recognition method can be used by feeding data to the system while recognizing students.

D. Objectives

The general objective of this project is to compare the efficiency of convolutional neural network and recurrent neural network in the application of facial recognition. The specific objectives are the following:

1) detect the faces of students through an image taken during a class;
2) recognize the students and record their attendance using the neural network being tested;
3) output result for the predicted students;
4) evaluate the speed and accuracy of the neural network used; and
5) compare the two neural networks used in terms of speed and accuracy

E. Scope and Limitations of the Study

The study will be conducted in a selected classroom during the second semester of the academic year 2019-2020 in University of the Philippines - Los Baños. The expected number of students to be tested will be at least five. The distance of the students from the camera that will be used will be dependent on the built-in face detection algorithm the camera is using.

II. Review of Related Literature

There are many researches on the usage of the integration of neural networks for face recognition, but there is no study yet on the comparison on the speed and accuracy of convolutional neural network and recurrent neural network for the usage of face recognition in attendance checking.

Students from Ankara University have developed a mobile system that records the attendance of students in their smartphones via face recognition [8]. The system uses the smartphones of its users and the face recognition process was also done using the smartphone. Eigenfaces, Fisherfaces, and Local Binary Pattern were used to develop an algorithm that will recognize the faces of the students. Results showed that the mobile system satisfies the needs of teachers, students, and parents in tracking the attendances.

Convolutional neural networks (CNN) consists of convolutional layers that contains different types of filters that will detect features from an object resulting in recognition of patterns and recognition of the object itself. In 2018, a study on face recognition via convolutional neural network on videos resolved the problem of severe image blurring, dramatic poses, and occlusion [9]. The study proposed a framework of convolutional neural network that will be fed still images and artificially blurred image to make the framework learn blurred insensitive images quickly. The framework was later called "Trunk-Branch Ensemble - Convolutional Neural Network" as it solves dramatic poses and occlusion by sharing the mid and lower convolutional layers between the trunk and branch networks.

In 2016, a joint convolutional neural network architecture was made to recognize the face and age of detected faces [10]. The architecture is called FDAR-Net or Face Detection and Attribute Recognition Network. The architecture combines three convolutional neural networks of feature-based recognition, specifically face detection, gender detection, and age detection. It detects all candidates of a face in a given image and determining every candidate if it is a face window or not. With 20 stages and a total of 26,000 face and non-face samples, the accuracy of FDAR-Net is 92.5% for the gender recognition and 95.4% for the age recognition with a margin of error of a year old.

Recurrent Neural Network (RNN) connects all nodes to one another and forming one direction. RNN is widely used in speech recognition and natural language processing however, in 2016, a recurrent neural network based architecture named Elman Levenberg Recurrent Neural Network was used together with a 3D Gray-Level Concurrent Matrix (GLCM) was used to achieve face recognition [11]. The 3D GLCM was used to extract characteristic features from images in the dataset used then were trained using Elman Levenberg Recurrent Neural Network. It achieved a 96.25% accuracy in recognizing faces. Moreover, emotion recognition was also achieved using recurrent neural networks by facial features [12]. The study determined five different emotions from a video feedback.

III. Materials and Methods

A. Overview

An application will be created to measure the speed and accuracy of Convolutional Neural Network and Recurrent Neural Network. The application will be created in Python 3.6.9 with the use of OpenCV, face_recognition api, Keras Deep Learning Library, and TensorFlow. The images that will be fed into the application will be captured using the Xiaomi Pocophone F1 with 12 megapixels camera.

B. Conduct of Study

1) *Class Selection*
   The class will be selected from a laboratory class in any computer science subject during the second semester of academic year 2019-2020. The students will be requested before hand that they will be used for the purpose of this study alone.
2) *Privacy Right*
   The students that will participate in this study will be confidential and their names will not be revealed. The students' names will be replaced with a label with the format: Person {number}.

C. Flowchart

The application will follow a concrete procedure in obtaining the input into giving the output needed for the study. It will give the face recognition results which will contain the labels of faces recognized and the accuracy of the recognition.

1) *Image Input*
   The images will be captured during the lecture hours of the selected class. Students will be asked to perform 10 different postures for the image analysis. The 10 different postures will be:

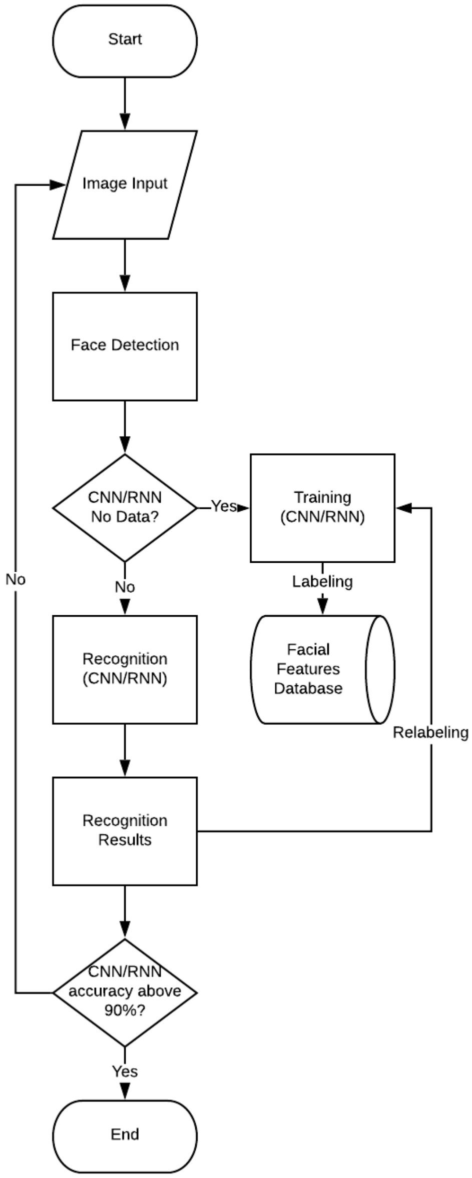

Fig. 1. Flowchart in Using Face Recognition via CNN/RNN

a) *Facing directly to the camera*
b) *Facing 30 degrees to the left*
c) *Facing 30 degrees to the right*
d) *Facing 30 degrees to the upward*
e) *Facing 30 degrees to the downward*
f) *Facing 30 degrees to the upper-right*
g) *Facing 30 degrees to the lower-right*
h) *Facing 30 degrees to the upper-left*
i) *Facing 30 degrees to the lower-left*
j) *Any angle determined by the subject*

2) *Face Detection*
To provide consistency of data, images that will be used in both frameworks for the neural networks will have the same image input and face detection algorithm as shown in Figure 1. The result of the detected faces will then be fed into the two different algorithms.
3) *Data Augmentation*
Datasets will have at most 10 data augmentation procedures for the dataset to fill in the insufficient amount of data input since having at least 100 pictures taken will consume time for the lecture hours.
4) *Training and Recognition*
The training and recognition process of the selected neural networks will be done consecutively with an adequate time interval to prevent decrease in performance due to heating. The training data that will be extracted during the run will be saved into the features database of the algorithm used.
5) *Convolutional Neural Network Model*
The convolutional neural network model that will be used will have three convolutional layers with 64 input cells, a window with size 3x3, activation function of rectified linear, and a pooling with size 2x2. The sizes of these layers are kept at a minimal size with a minimum amount of data received. The final activation function that will be used is softmax to receive outputs of percentage predictions of classes.
6) *Recurrent Neural Network Model*
The recurrent neural network model that will be used will start with two Long-Short Term Memory layer with 64 input cells followed with a dropout layer of 30%. The size of the layers is to keep a minimum size for a minimum amount of data received and to match the convolutional neural network model. The final activation function that will be used is softmax to receive outputs of percentage predictions of classes.
7) *Recognition Results*
The results will contain the predicted label of the face being recognized and the accuracy of the recognition. The results will be saved in a log where the history of the results will be seen. If the results shows inaccurate assumptions of the faces, relabeling will be done. The differences in accuracy and speed will be determined on the training and recognition capabilities of the selected neural network.

D. *Analysis of Results*

The analysis will be performed after both types of neural network will achieve a 90% percent accuracy or more. The results will be analyzed and interpreted by measuring the percent accuracy of the neural network in recognizing the faces of the input. It will also determine the minimum number of inputs required to get an accuracy of at least 90%.

IV. Results and Discussion

A. *Loading Dataset*

Input images are taken as a class and individual faces are extracted as shown in Figure 2. The individual faces are then labeled to their corresponding classes. Data augmentation is then used to produce more data to feed the neural network as shown in Figure 3. The number of images after data augmentation is always 100. 10% of the data-set will be used as testing data thus will not be used for training. The attendance of the classes is recorded manually at this time.

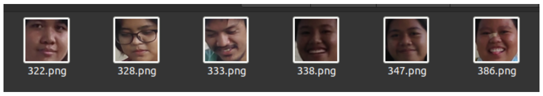

Fig. 2. Sample Image Inputs

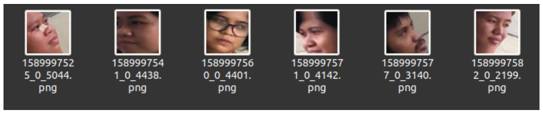

Fig. 3. Sample Data Augmented Image

B. Training Datasets

After loading the data set, neural network training is implemented by training using the data set per day using the model of the neural network used as described. The loss, accuracy, val loss, and val accuracy is recorded per data set inputted in the model. As shown in Figures 4 and 6, the CNN model's accuracy and val accuracy is higher than that of the RNN model's by a huge margin. And in Figures 5 and 7, the CNN model's results are better than that of the RNN model's.

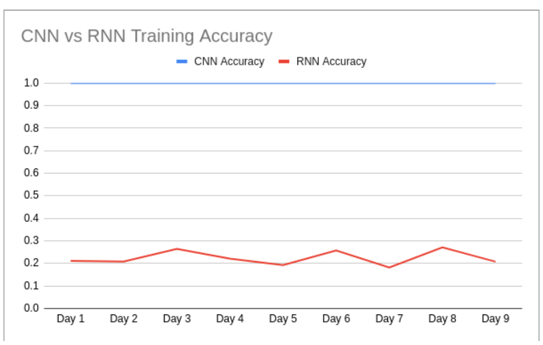

Fig. 4. CNN vs RNN Training Accuracy

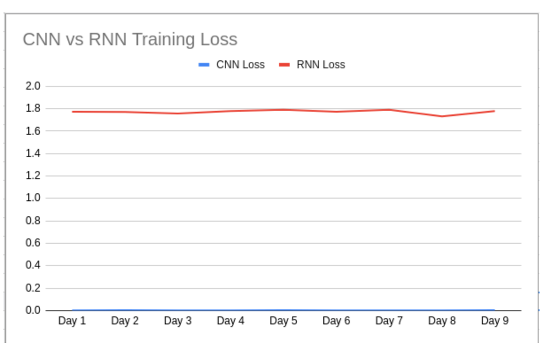

Fig. 5. CNN vs RNN Training Loss

C. Testing Neural Network Models

After every training of the neural network used, it is used to predict the testing data separated from the data set to measure

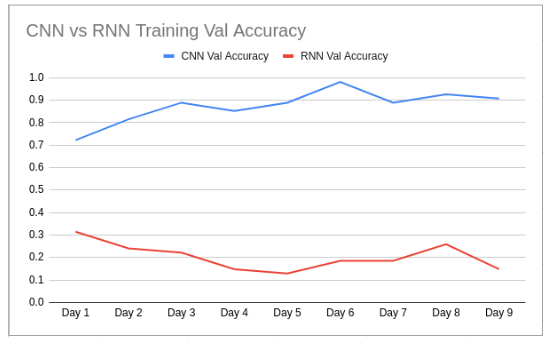

Fig. 6. CNN vs RNN Training Val Accuracy

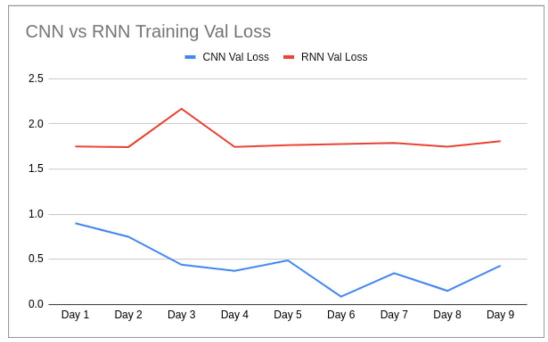

Fig. 7. CNN vs RNN Training Val Loss

the testing accuracy of the current model. As shown in Figure 8, only the CNN model achieved a consistent 90% overall accuracy starting from day 6. Furthermore, the RNN model's accuracy is fluctuating from all days of the data set used. The fluctuation of the RNN model's accuracy can also be seen on the individual classes' accuracy as shown in Figures 9, 10, 11, 12, 13, and 14 compared to the CNN model's individual classes' accuracy which is more consistent in predicting the classes. It can also be seen that the RNN model is not feasible for this use-case scenario as it predicted only one class for all classes on the last days of the data set.

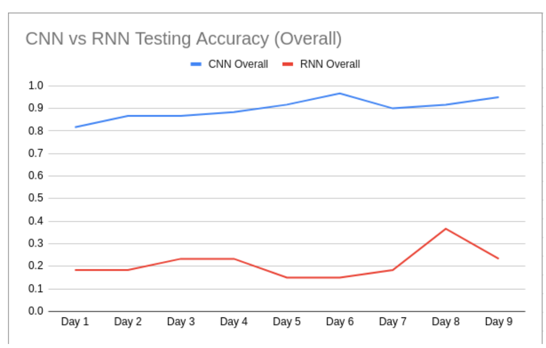

Fig. 8. CNN vs RNN Overall Testing Results

D. Overall Feasibility

The CNN model can achieve a consistent accuracy of at least 90% starting from Day 6 while the RNN model cannot achieve the minimum requirement of 90% based on the limited

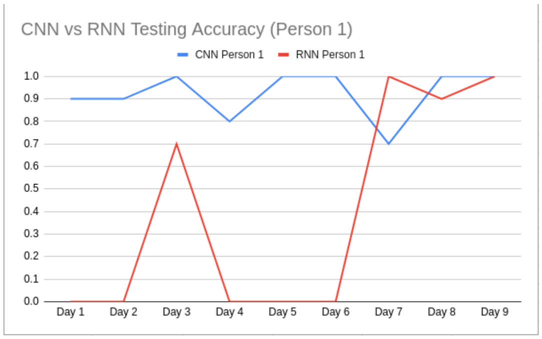

Fig. 9. CNN vs RNN Person 1 Testing Results

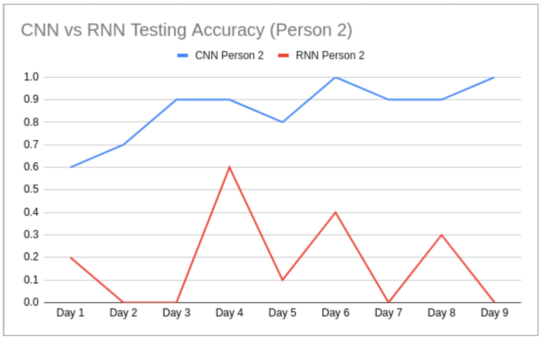

Fig. 10. CNN vs RNN Person 2 Testing Results

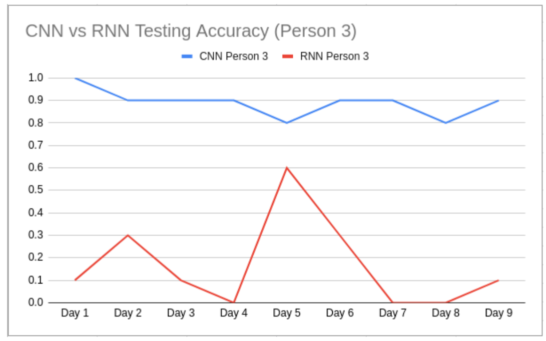

Fig. 11. CNN vs RNN Person 3 Testing Results

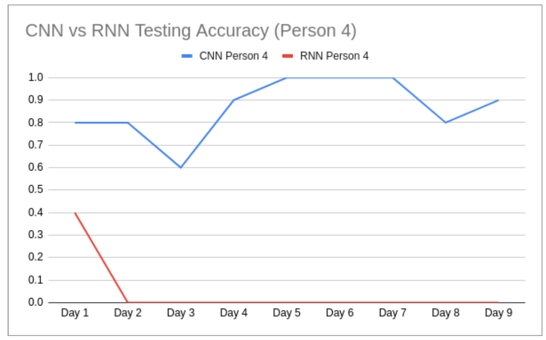

Fig. 12. CNN vs RNN Person 4 Testing Results

data set collected for this study. Thus, only the CNN model is feasible for this type of scenario.

V. Conclusion

This study shows that the CNN model can be used for the recording of the attendance of a class given that the first six days of lecture may not give proper results of the record of students thus the possibility of skipping the record for attendance on the said days. The study also shows that the CNN model is clearly better in image classification compared to the RNN model.

VI. Recommendation

Further study of different models that can be used for image classification can be conducted to widen the range of feasible neural network models that can be used in this use-case scenario. Furthermore, the mix of different number of layers, node inputs, pooling, and such can be modified to further improve the CNN model used. For data collection, more pictures can be taken into account to further improve the accuracy of the models used. To further maximize data collection, video input can be used to have multiple data sets and replace data augmentation.

References

[1] V. Kassarnig, A. Bjerre-Nielsen, E. Mones, S. Lehmann, and D. D. Lassen, "Class attendance, peer similarity, and academic performance in a large field study," *PLoS ONE*, vol. 12, no. 11, nov 2017.

[2] P. K. Pani and P. Kishore, "Absenteeism and performance in a quantitative module A quantile regression analysis," *Journal of Applied Research in Higher Education*, vol. 8, no. 3, pp. 376–389, jul 2016.

[3] S. S. Bhairannawar, S. Sarkar, K. B. Raja, and K. R. Venugopal, "Implementation of Fingerprint Based Biometric System Using Optimized 5/3 DWT Architecture and Modified CORDIC Based FFT," *Circuits, Systems, and Signal Processing*, vol. 37, no. 1, pp. 342–366, jan 2018.

[4] F. M. Al-Naima and H. A. Ameen, "Design of an RFID based students/employee attendance system, Tech. Rep. 1, 2016.

[5] D. Thakkar, "Biometric Devices: Cost, Types and Comparative Analysis," 2018. [Online]. Available: https://www.bayometric.com/biometric-devices-cost/

[6] B. Ray, "A Breakdown Of 7 RFID Costs, From Hardware To Implementation," 2018. [Online]. Available: https://www.airfinder.com/blog/rfid-cost

[7] A. Alazzawi, O. N. Ucan, and O. Bayat, "Performance of Face Recognition System Using Gradient Laplacian Operators and New Features Extraction Method Based on Linear Regression Slope," *Mathematical Problems in Engineering*, vol. 2018, 2018.

[8] R. Samet and M. Tanriverdi, "Face recognition-based mobile automatic classroom attendance management system," in *Proceedings - 2017 International Conference on Cyberworlds, CW 2017 - in cooperation with: Eurographics Association International Federation for Information Processing ACM SIGGRAPH*, vol. 2017-Janua. Institute of Electrical and Electronics Engineers Inc., nov 2017, pp. 253–256.

[9] C. Ding and D. Tao, "Trunk-Branch Ensemble Convolutional Neural Networks for Video-Based Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 1002–1014, apr 2018.

[10] H. Liu, X. Shen, and H. Ren, "FDAR-Net: Joint convolutional neural networks for face detection and attribute recognition," *Proceedings - 2016 9th International Symposium on Computational Intelligence and Design, ISCID 2016*, vol. 2, pp. 184–187, 2016.

[11] R. Y. Dillak, S. Dana, and M. Beily, "Face Recognition using 3D GLCM and Elman Levenberg Recurrent Neural Network," Tech. Rep., 2018.

[12] A. Mostafa, M. I. Khalil, and H. Abbas, *Emotion Recognition by Facial Features using Recurrent Neural Networks*, 2018.

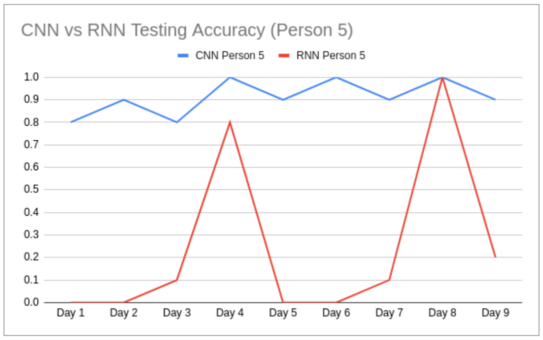

Fig. 13. CNN vs RNN Person 5 Testing Results

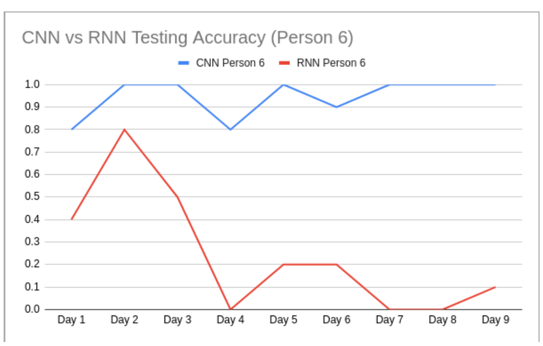

Fig. 14. CNN vs RNN Person 6 Testing Results