Predicting Paper Breaks in a Paper Machine Using CNN's

Using Image Classification to Predict Impending Process Failures in a Real-World Dataset

This summarizes the use of CNN's to predict an impending process failure in a paper machine using a real world dataset in this Github repo. This is an extension of a prototype trial where I created a synthetic dataset and trained a CNN image classification model on images representing normal conditions and pre-failure time slices as a warning state. I published that in an article in Towards Data Science (reference below). I cannot publish this work in TDS because the data license does not allow for commercial use. By extension, my work is shared per the dataset license (see end of article) but commercial use is prohibited.

CNN's to Predict Process Failures
A Prototype — Stack Multiple Time Slices in a Single Image and Classify: Normal, Warning or Failure
towardsdatascience.com

The data I used is available from (Chitta Ranjan, Markku Mustonen, Kamran Paynabar, Karim Pourak, "Rare Event Classification in Multivariate Time Series", arXiv:1809.10717, 2018 (the license repeated at the end of this summary and in my Github repo, commercial use is prohibited).

Introduction

The goal is to arrange data from several time-slices into an image that is a snapshot of past and current state. If conditions prior to a failure are different from normal operation, then images of the condition can warn of an impending failure. The purpose isn't to classify a failure vs. a normal operation, but to classify a running state as a warning of an 'impending failure'.

Why do this?

Industrial systems run on process controls. These keep the system running within parameters. Yet, systems shut down. Excluding other factors such as manual shut down, or equipment failure, upsets still happen. Identifying an impending failure could give operations staff time to prevent an upset. Additionally, analyzing a process in this way may provide insights into why or what combination of events lead to an upset and provide an opportunity to adjust the control system.

Dataset

This dataset covers approximately one month of operation of a paper machine with samples recorded at two-minute intervals. The features are not identified with specific process variable names and the data may have been adjusted for IP purposes. There is one categorical feature that appears to split the data into eight discrete operating scenarios which may be related to the type or weight of paper or some other parameter. I selected the most frequent subset making up 36% of the data from this feature (where feature X28 is 96) and reduced the number of features via data analysis. Knowledge of the specific features could improve this step.

The paper breaks and feature X28 are shown below. The paper breaks are the vertical lines in time. The blue line is feature X28. The value 96 occurs in the beginning, middle, and end of the month as indicated by the arrows.

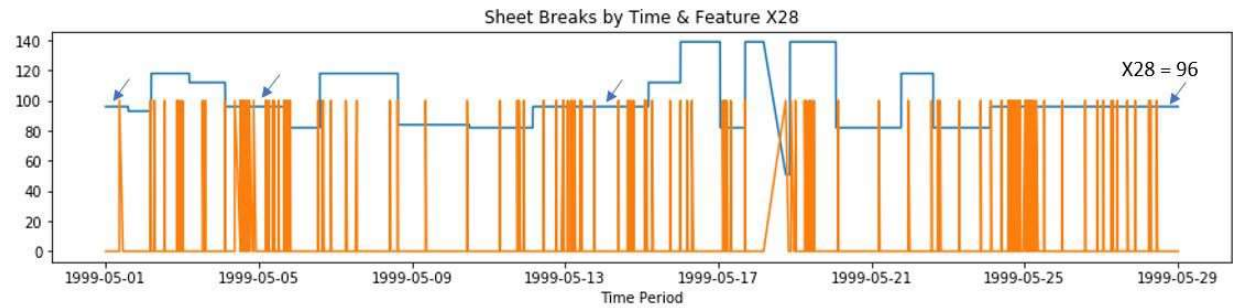


Image by Author

More data and knowledge of the meaning behind each feature may produce a better model that could also be refined with new data month over month.

This dataset only has process values recorded every two minutes. The original raw data is probably more detailed in time. Different or multiple time intervals could be used with other datasets to gain insights or provide short or longer period predictions.

Converting Data to Images

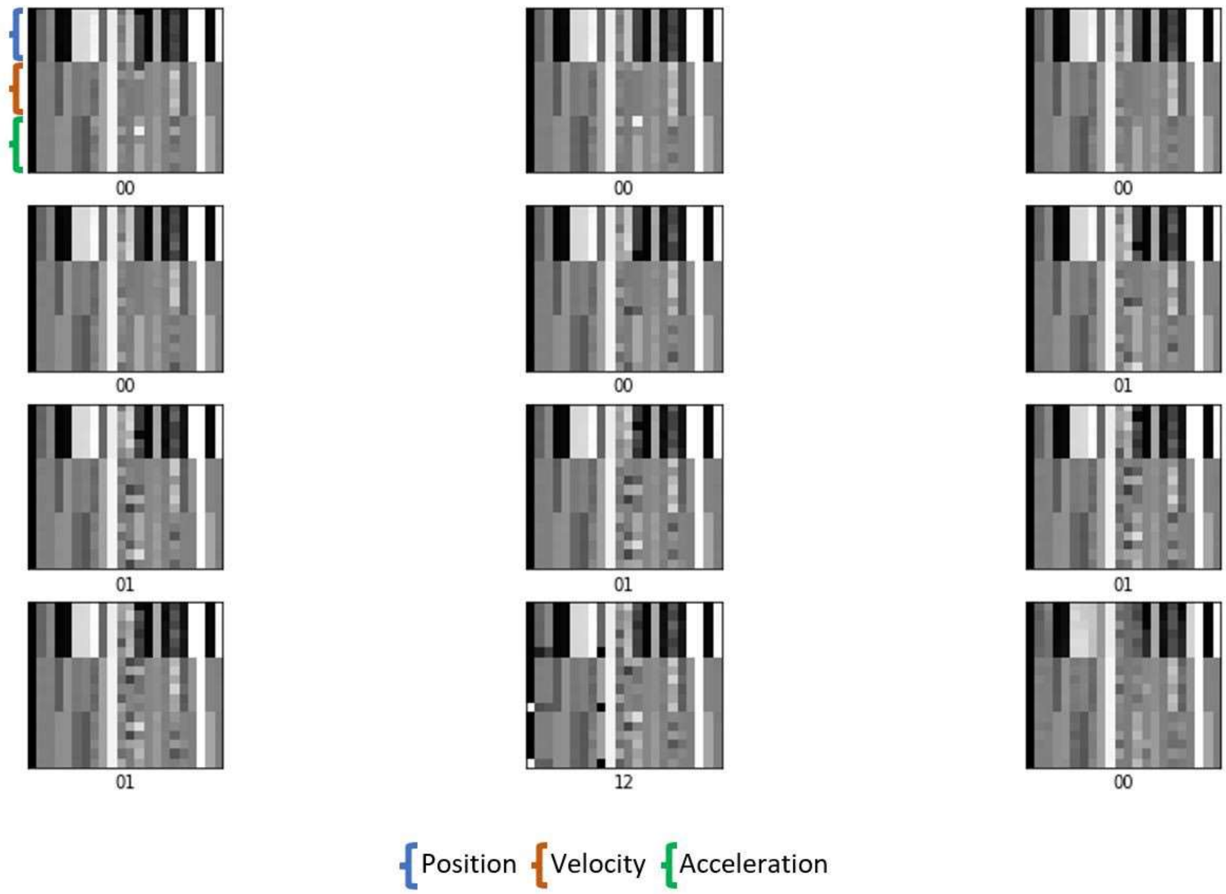
The data is relabelled from 0 and 1 for normal and paper break, to 0 for normal, 1 for warning, and 2 for paper break. The five time samples preceding a paper break are set to label 1.

The data is then scaled on a 0 to 1 range and 6 periods of data are assigned to an image covering a 12 minute period. Each successive image overlaps the previous one by 5 time periods. Each image is categorized by the label corresponding to the last time slice in that image: normal, warning, or paper break.

The images are built in 3 layers with the process variable values or 'position' as the first layer. The derivative of the process values with time (like velocity) is the second layer, and the derivative of the 'velocity' layer is added as an 'acceleration' layer.

Once the failure time slice is in the last row of an image, the process repeats by building the next image starting with the next 6 time periods following the previous failure time slice.

The picture below shows twelve consecutive images built from the dataset leading up to a paper break and the last image representing the condition after restarting the machine. The control point value (position), ‘velocity’, and ‘acceleration’ bands are visible in each image. The image order is left to right by row with the upper left as the first image and lower right as the last. The numbers on the left are the original normal (0) and paper break (1) labels corresponding to the last time sample in each image. The numbers on the right are the adjusted labels with normal (0), warning (1), and paper break (2). The first five images are normal operating (00). The next five are warning condition (01) of the five samples prior to the actual paper break. The paper break is next (12) which is followed by a normal restart (00). All 6168 samples are created this way.



5 normal images, then 5 warning images, the paper break image, and a restart image
Image by Author

Training

Training is similar to the prototype CNN model. I used a Sequential CNN model with two *Conv2D*, two *MaxPooling*, and a *Flatten layer* feeding a *three-layer* 128–64–3 network with *Dropout*. I did not undersample the normal class to keep as much variability in what a normal condition 'looks like'. Thus the high initial accuracy. Details can be found in the Jupyter Notebook in this repository "Paper Machine Sheet Break Subset96-3Class w Dates Adamax"

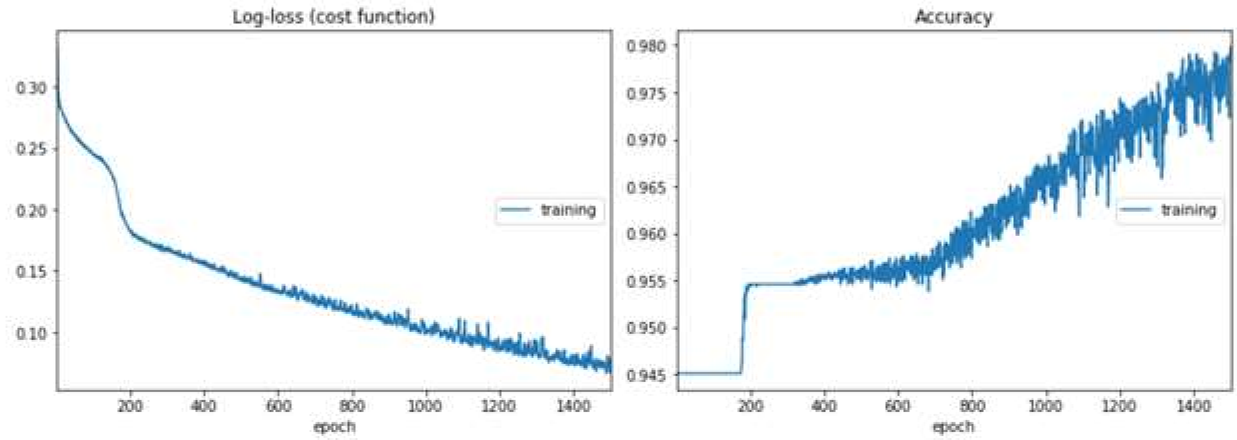


Image by Author

Results

I applied a custom classifier to the model class probability predictions (predict_proba) rather than using the standard argmax prediction due to the imbalance in the dataset. Most of the images, about 94%, are from the normal class, and warning and paper break images make up about 5% and 1% respectively. The custom classifier eliminates a class if the predicted class probability is below the frequency of that class in the dataset. For example, if a warning and paper break probabilities are 8% and 3%, then the normal probability would be 89% and then is excluded. The prediction is the argmax of the remaining two probabilities and would be a warning.

This gives better prediction of warning conditions than the standard prediction as shown in the two following test set confusion matrices. The first is the custom classifier and the second is the standard prediction.

```
x_test Confusion Matrix - Adjusted Prediction

array([[1793,  248,    0],
       [  39,   59,    0],
       [   0,    0,   20]], dtype=int64)
```

```
x_test Confusion Matrix - ArgMax

array([[1995,   46,    0],
       [  78,   20,    0],
       [   0,    0,   20]], dtype=int64)
```

Image by Author

The custom classifier has a higher true positive rate on the warning state, but also a higher false positive rate predicting a warning state versus a normal state about 12% of the time. The standard classifier only has a 20% success rate in picking warning states but has a low 2% rate of labelling normal conditions as a warning condition.

The overall accuracy of finding warning states in the test data set was about 60% of those labelled as warning states. In some cases, not all five intervals prior to a failure in the test set were classified as warnings. However, most failures were preceded by warnings.

When looking at the results from the confusion matrix and the charts below, I see three things could be happening with this model and dataset. Firstly, the model just may not be performing very well.

Secondly, there are many samples where the process is similar to a pre-failure condition and some of the false positives for a warning state are actually times when the system is close to having a paper break. This would be insightful, but cannot be verified with the information I have.

Thirdly, there may be a bias built into the way I labelled the warning states. I picked 5 time periods (10 minutes) before an actual break to be labelled as a warning to provide some practicality for time to intervene in the process. Conditions in the paper machine may develop to a paper break much faster, and thus 2 or 3 of the warning states before some, or all of the breaks, may not be representative of a process deviation. Thus labelling data as a warning state when it is perfectly normal could lead the model to predict small process deviations as evidence of an imminent paper break and / or reduce the performance of predicting warnings in the test set.

Reducing the number of warning states before a paper break may improve the model but would also reduce the time to intervene to prevent a break.

Charts

I have included several charts to give an overview of how the model performed as well as how different periods in the data appear as a smoothly running machine or as a somewhat unstable machine.

The charts show the test and training results (running the train set through the final model) plotted together by the day and time of the samples. I present them this way because plots of the training and test sets ordered consecutively by sample don't show how performance changes through the month because the samples are shuffled. Also, separate plots of the test and training sets by time have gaps from the missing data in either the test or training set.

These plots have two tracks. The upper track has the predicted class, normal (0), warning (1), and paper break (2) from the model. This track also has the actual labels for warnings as circles, and the paper breaks as diamonds. The test predictions are blue dots, and the training predictions are blue X's. With shorter time intervals, it is easy to see where the prediction and actuals align.

The lower track has the probability of class (0 to 1) plotted for test and training predictions. Here test results are dots and training results are X's. The probabilities by class are green for normal, yellow for warning, and red for paper breaks (also labelled as faults).

Chart 1 shows the distribution of the sample subset through the month with 58 paper breaks in this data subset.

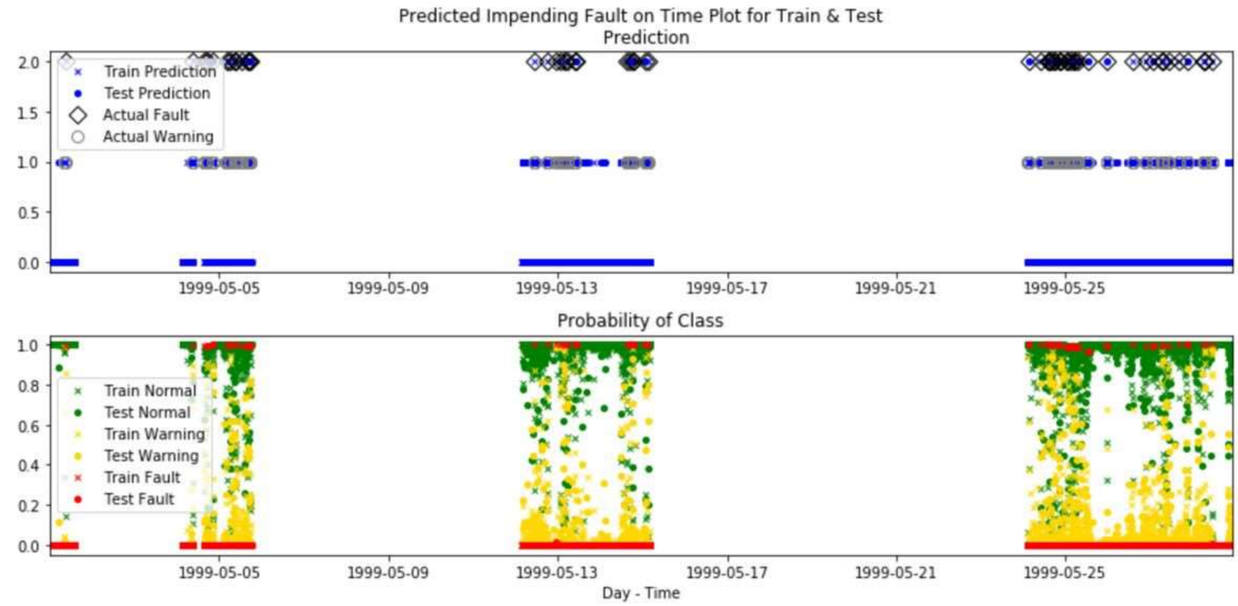


Chart 1 — Image by Author

Chart 2 is a 4–1/2 hour running period on May 5 from 13:00 to 17:30 hours. This shows how a small drop in normal probability below 94% is accompanied by a warning probability above its occurrence frequency in the data and thus a warning prediction. There are several 'false' warnings which could be due to irregular operating conditions. These periods of high 'false' warnings tend to happen more often close to a paper break or when there is a short run time from start-up to a paper break. The correct warning and paper break predictions are visibly aligned with the actual markers.

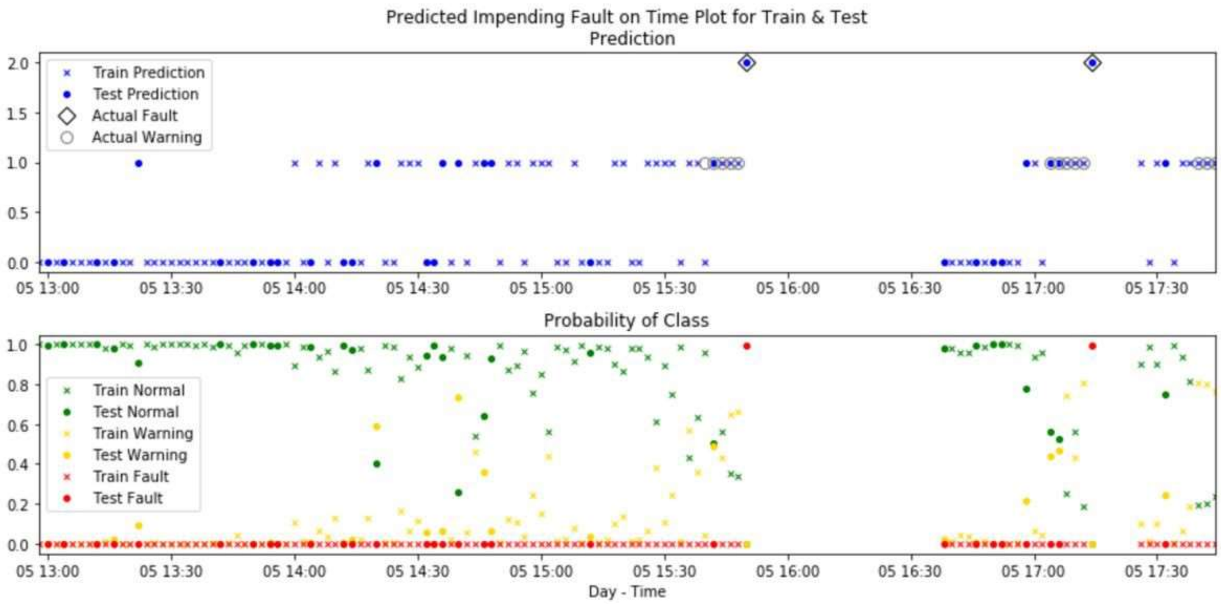


Chart 2 — Image by Author

Chart 3 has a gap where 68 minutes of missing data. It appears that the machine ran continuously until the paper break at 23:00 hours. This interval has far fewer false warnings than in Chart 2 above.

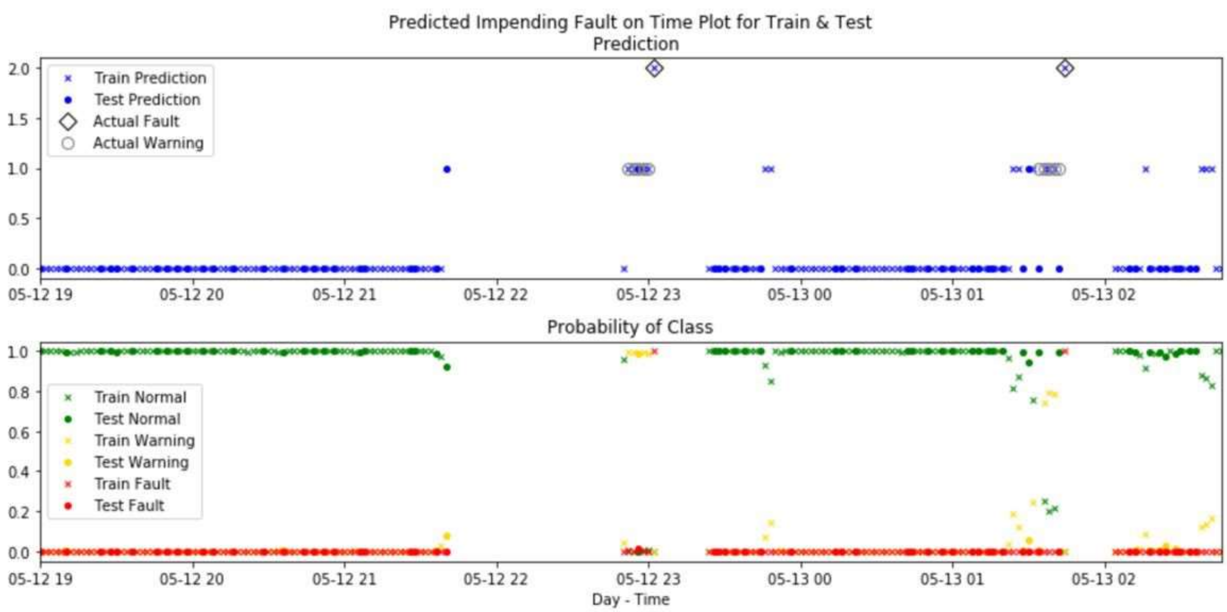


Chart 3 — Image by Author

Chart 4 shows a smooth run time over about 28 hours. The false warnings are more frequent after the machine restarts (on the left) and before the paper break at about 15:00 hours on the 14th of the month.

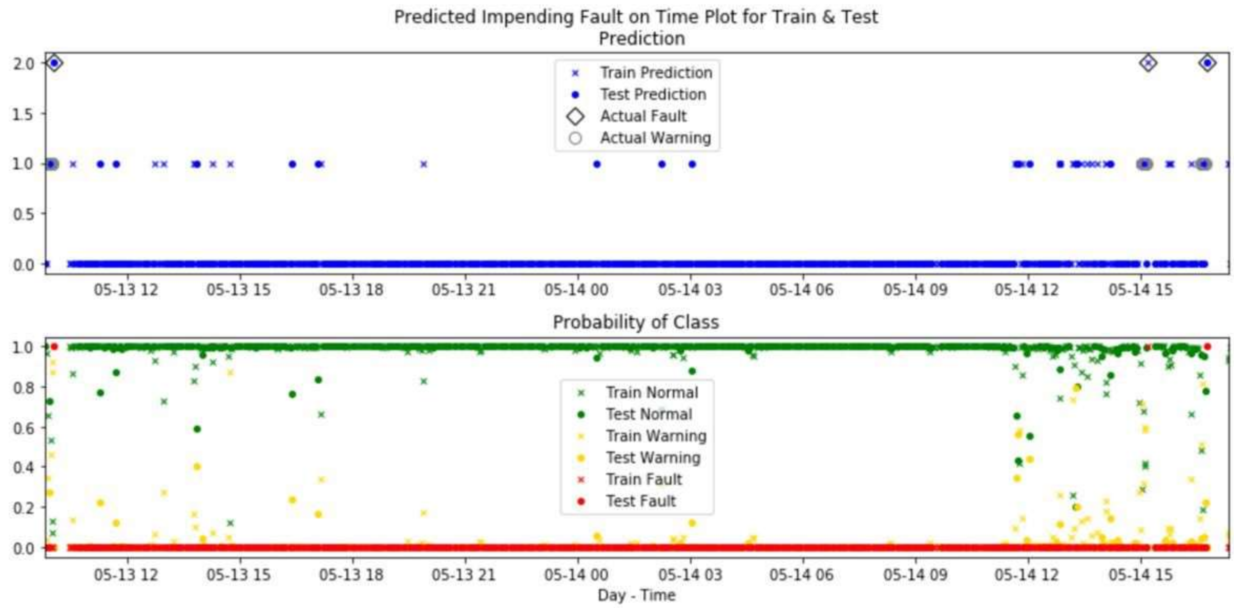


Chart 4 — Image by Author

Chart 5 shows a very clean 11 hour run time that precedes Chart 6.

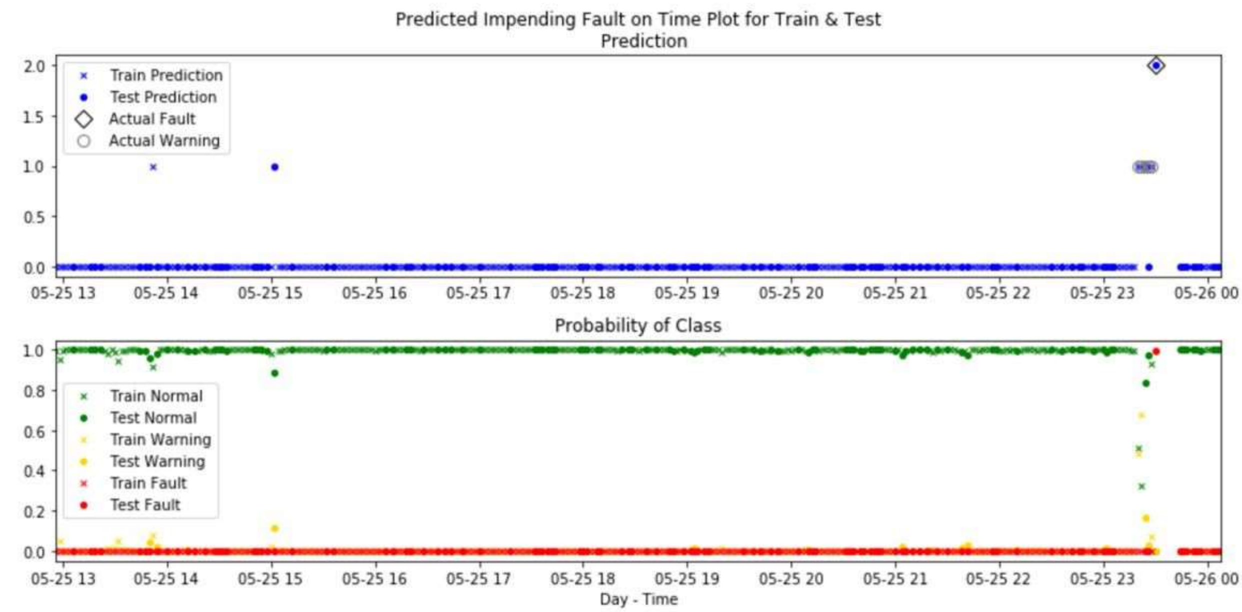


Chart 5 — Image by Author

Chart 6 follows Chart 5 above in time and is another example of increasing number of false warnings prior to a paper break.

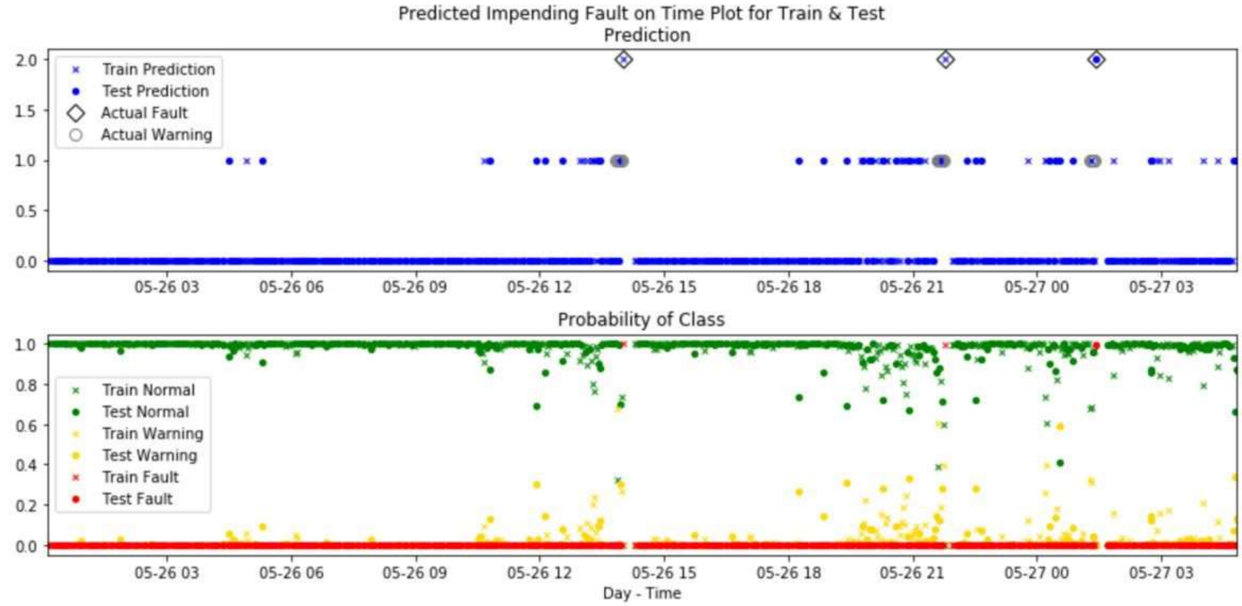


Chart 6 — Image by Author

Recap

This dataset is almost twenty years old, and the paper machine and other similar ones are likely much different now, perhaps with more sophisticated control systems. This approach could be used in other processes, not necessarily as a control system, but to explore upsets and fine tune systems.

The objective was not to classify paper breaks versus normal operating conditions since identifying the paper break after it has happened doesn't have any operational benefit. Identifying changes in operating conditions that are likely to lead to a paper break potentially has value for operator intervention, improving process control logic, or understand second order or compounding effects of concurrent shifts in multiple process measurements.

This work showed that a CNN model could be used to predict or warn of an impending process failure on a real-world dataset. The model could be improved with more data and knowledge about the actual paper machine that produced the data.

I welcome your comments.

Use or application of methods included here are at the sole risk of the reader. The author takes no responsibility for application by the reader.

The dataset license from the dataset owner is shown below.