Temporal Modeling Recommendations

Project: Temporal Deep Learning for Automated Grading of Retinal Inflammation in FA

Tymo van Rijn – 1057297

Last updated: 3 Mar 2026

---

1 Scope of This Advice Note

This short advisory note summarises the key findings from Week 1, Week 2 and Week 3 analyses and the first sequential experiments. It supports supervisors and future project members in deciding *which temporal strategy* to prioritise during the remaining development weeks.

2 Summary of Evidence

Evidence	**Key insight**
Week 1 analysis presentation (PDF)	Later frames carry more discriminative signal; naïve averaging improves *window defect* & pooling but can dilute staining information. There also is a huge variability in number of frames, with big outliers. The intensity over time probably won't give us a lot of information for classifying the different HyperF_Types. Clinicians show in literature that information lies in temporal behavior, not just static intensity.
Week 2 timestamp analysis (PDF+CSV)	85 % of frames lie in the late (>60 s) phase; exam durations span 1–30 min; temporal coverage is highly unbalanced. These elapsed times do not correspond with the actual FA phases clinicians use, so binning them in that way would not make sense. The *(new) last frame only* baseline outperforms the *last 2 frames averaged* baseline. Binning is used across different other studies, and seemed to get good results

Week 3 LSTM binning results	The LSTM results showed an improvement compared to the *last frame* baseline, using the binning strategy. Number of binning influences the results , higher n_bins seems to result in better model performance. The binning strategy being used also influences the results, for now the strategy where "logic" is applied, performs best.

3 Recommendation Matrix

Option	**Effort**	**Expected gain**	**Risks**	**Recommendation**
Phase-aware bins (Choroidal → Late)	Medium	Medium	Phase cut-offs uncertain	Investigate if timestamp OCR stabilises & adopt if robust
Full-sequence **LSTM** ($\geq$ 10 frames)	Medium	Potentially high	Overfit, variable-length padding	Since higher n_bins seems to help, this would be a logical next step
Using **GRU** instead of **LSTM**	Medium	Low	Same results	Might be waste of time, since principle is the same

4 Concrete Next Steps (4 weeks)

1. **Finalize bin implementation**
   - Implementing an elapsed time OCR-extracted bins approach.
   - Run some more tests using different n_bins.
2. **Implement full-sequence LSTM**
   - This increases the number of frames used, could include more "blur", but could also benefit from more temporal data.
   - Probably excluding exams that have a very high amount of frames, will have to find out what is an appropriate n_frames by using my already existing analysis for *frame distribution*.