BAITRADAR: A MULTI-MODEL CLICKBAIT DETECTION ALGORITHM USING DEEP LEARNING

Bhanuka Gamage, Adnan Labib, Aisha Joomun, Chern Hong Lim, and KokSheik Wong

MONASH University

1129

1. Introduction

- YouTube has gained popularity among multi-billion dollar company marketers.
- YouTubers exploit this opportunity to make money: *advertisements* ⇒ Offer money for each view where the video being compliant with stipulated advertisement-friendly guidelines.
- Leads to ***clickbait*** ⇒ provokes users to click on videos using attractive titles and thumbnails.
- 😞 users end up watching a video that does not have the content as publicized in the title.
- Why? Humans have the need to explore and fill the knowledge gap between *what they know* and *what they wish to know* [Information gap theory].
- Clickbait exploits this gap by using the Title and Thumbnail.

Importance of solving the Clickbait problem (motivations):

(a) Reduces time wasted online and unnecessary network traffic.
(b) Safer environment for children to navigate the internet.
(c) Recommendation engines on social media platforms can be improved to not recommend clickbait.
(d) Improve the overall experience for Social Networking Services (SNS) platform users.

2. Research Gap

- Clickbait detection for web pages and articles - available and well researched.
- Video clickbait detection - only a handful of research is available.
- No implementation of a model with an architecture with **audio transcript** and other important cues such as titles, tags etc. are taken into account.

3. Research Purpose

✓ Address the Clickbait video issue on YouTube ⇒ develop an algorithm (*BaitRadar*) using a multi-model deep learning architecture.

- Effectively and efficiently detect Clickbait on YouTube.

4. Significance of Study

The BaitRadar algorithm:

- Combines **multiple cues** in a YouTube video to make inferences.
- Exploits the **audio transcript** of the video in tandem with other cues in the video.
- Achieves **more resilience** against missing data ⇒ a more robust classification approach against the conventional methods.

5. Data and Methodology

Dataset

8591 clickbait videos

5049 non-clickbait videos

Training (81%)	Validation (9%)	Testing (10%)
11340	1260	1400

Methodology

✓ Proposed architecture - integrates 6 different deep learning models, each looking into different attributes of a clickbait video.
✓ Individual models ⇒ trained to perform on attributes of the video and provide a prediction.
✓ Multi-model architecture⇒ fuses all prediction outcomes to infer whether the video of interest is clickbait or not.

The following steps are performed to develop our architecture:

1. Dataset gathering and pre-processing.
2. Training and evaluation of individual models.
3. Multi-model deep learning architecture.

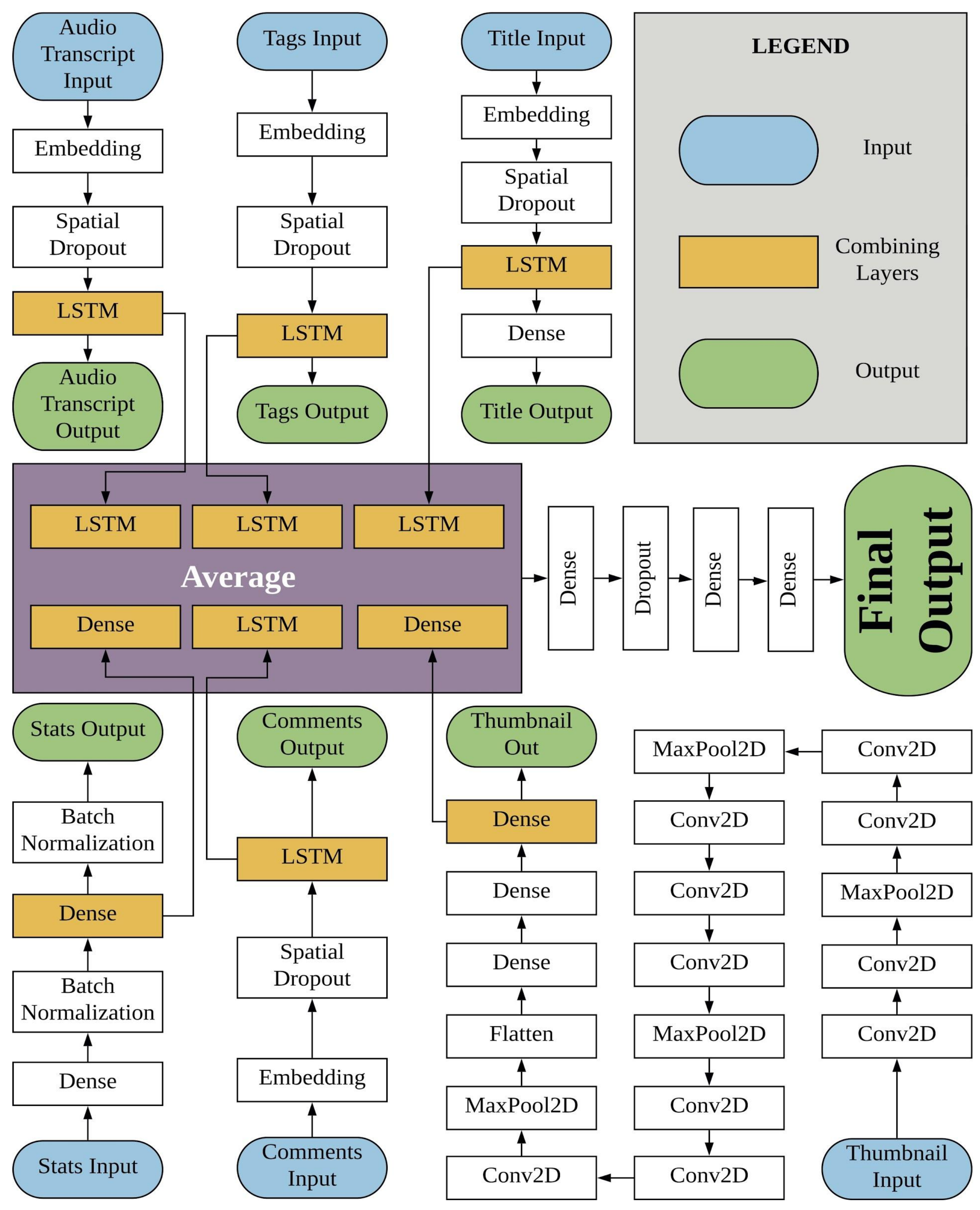

Fig. 1: The final combined architecture of our model

6. Experiments and Findings

- Model implementation: Tensorflow with Keras using Python and trained on Nvidia RTX 2080 Ti.

Table 1: Detection accuracy from individual models

Model	Training Accuracy	Validation Accuracy	Testing Accuracy
Audio Transcript	99.09	94.12	93.80
Title	94.29	87.20	87.30
Thumbnail	90.61	90.61	81.43
Comments	98.64	96.40	96.80
Tags	99.83	98.76	98.70
Statistics	78.69	78.69	77.78

- Highest individual accuracy: Tags model ⇒ Reason? As clickbait videos contain higher number of tags for tricking the YouTube recommendation algorithm.
- Lowest individual accuracy: Statistics model ⇒ Reason? Clickbait videos not having easily identifiable outstanding statistics.

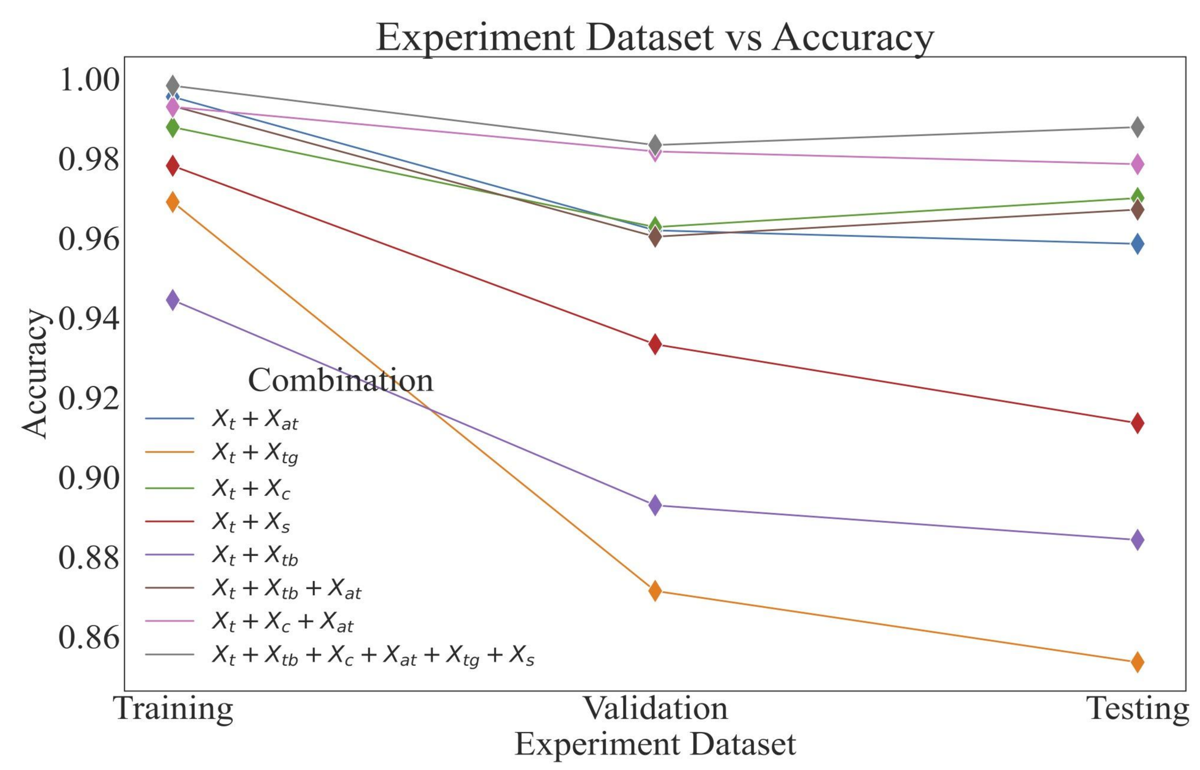

Fig. 2: Summarised results for different combinations

- Combination of 6 models: Highest accuracy (~98%) with an inference time of ≤ 2s ⇒ Reason? It has the most data to train and infer with = *BaitRadar.*
- Combining all models ⇒ takes more epochs on average to converge against other models. Reason? Because each (sub-)model is contributing to the overall accuracy.

7. Limitation and Conclusion

- Issues with borderline cases ⇒ videos that are not clearly clickbait: from the perspective of a subscriber and non-subscriber.

✓ Given the results and analysis ⇒ BaitRadar performs as intended.

✓ Future work: explore the performance of the model on various kinds of media ⇒ aiming to understand different insights through model interpretability to better detect clickbait videos.