List of gaps:

* **META MODEL:**
  + Optimize model fusion techniques (averaging, stacking, or weighted combination)
* **ATTENTION MODEL:**
  + Frame level attention
  + Temporal attention
* **Mesonet/Meso Inception net:**
  + These are vulnerable to adversarial attacks, where small perturbations to the input data can cause misclassification.
* **Multi Attentional model:**
  + It is unclear how the model performs when there are changes in the lighting conditions or camera angles in the input videos.
* **Convolution vision transformer:**
  + ViTs require large amounts of training data to achieve state-of-the-art performance, this is a big drawback since we don't get much data on deepfakes
  + ViTs are sensitive to the resolution of the input images. The model's performance can deteriorate significantly when the input images have low resolution
  + Tested on only 400 videos, classifies the non-face region part of the image wrongly,loss value is 0.32
* **Unsupervised learning(Clustering):**
  + Can detect only if the face is completely visible, less accuracy on Low Quality videos dataset
  + This paper has very less analysis on feature extraction and hyperparameters, that's why the accuracy on low quality images is low
* **Hybrid CNN model:**
  + The model uses only one attention mechanism in the proposed hybrid model,so it needs proper orientation in order to map face landmarks
  + ResNet v2 is prone to overfitting, especially when working with small datasets, in this paper they have trained and tested on less than 4000 videos and not even tested on different datasets,so the model may be overfitting
  + Both the techiniques used in hybrid model doesn't work well on low quality images, and there's no mention about low quality image data in the paper
  + If any obstruction is present in the way of the illumination and the camera then the model fails.
  + If the person is seeing towards any other angle other than the direct camera angle then the model fails.
* **Multiple Instance Learning:**
  + One of the main challenges of MIL is that it can be difficult to interpret the results. This is because MIL works with bags of instances rather than individual instances, making it challenging to identify which instances are most critical for a particular classification decision.
  + Gradient vanishing problem