

DC Project

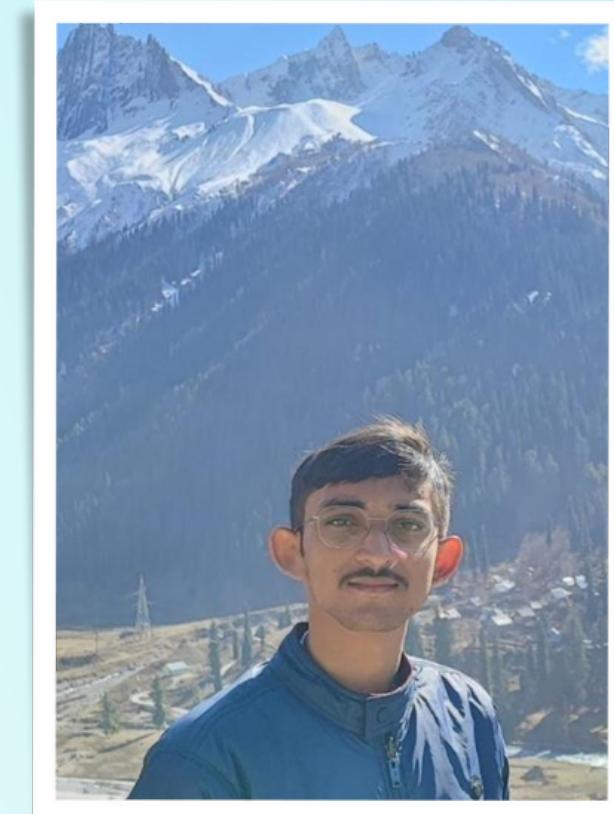
VISION TRANSFORMER ACCELERATOR (SOFTWARE)



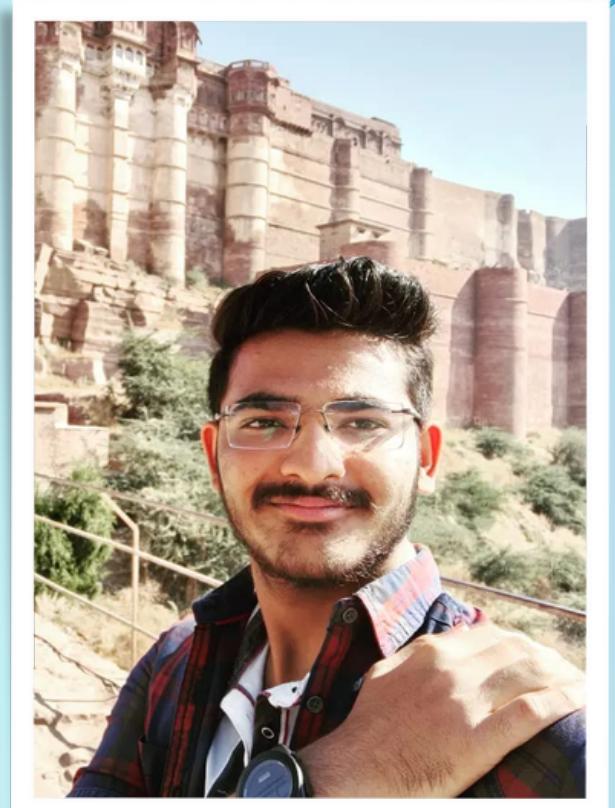
OUR TEAM



DR. PALASH DAS
MENTOR



VATSAL DADHANIYA
STUDENT

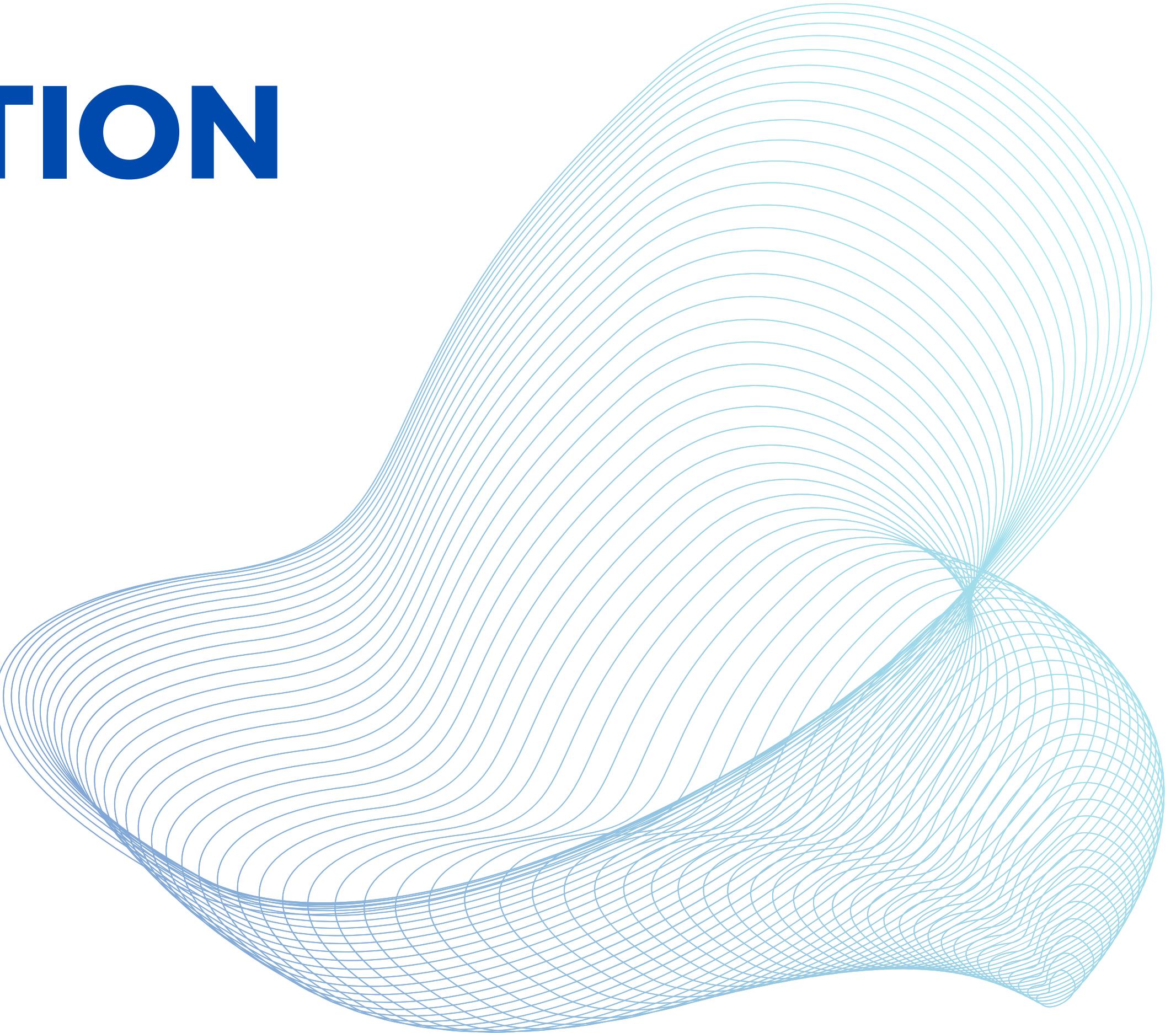


SAGAR VEKARIYA
STUDENT

INTRODUCTION

Our aim is to reduce the no of calculations in vision transformer model by removing the unnecessary multiplication of same pairs(input and weights) in a matrix.

To increase the chances of finding same pairs we will round off all the weights and corresponding inputs to a particular decimal place. Doing so will result in increase of percentage hits(percentage of same pairs) but will also reduce the accuracy of model.



PREREQUISITE WORK

1

Machine Learning

We saw the 3 courses of Andrew Ng related to basic understanding of supervised learning.

2

Deep Learning

Saw a 24 hour lecture of Daniel Brouke to learn how to code deep learning models.

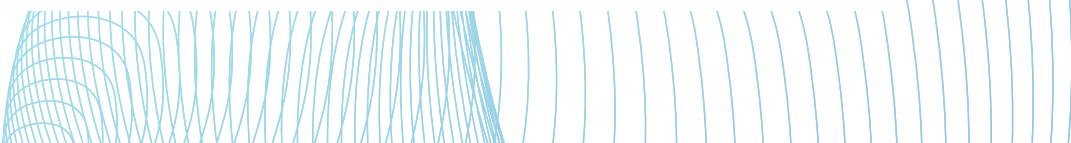
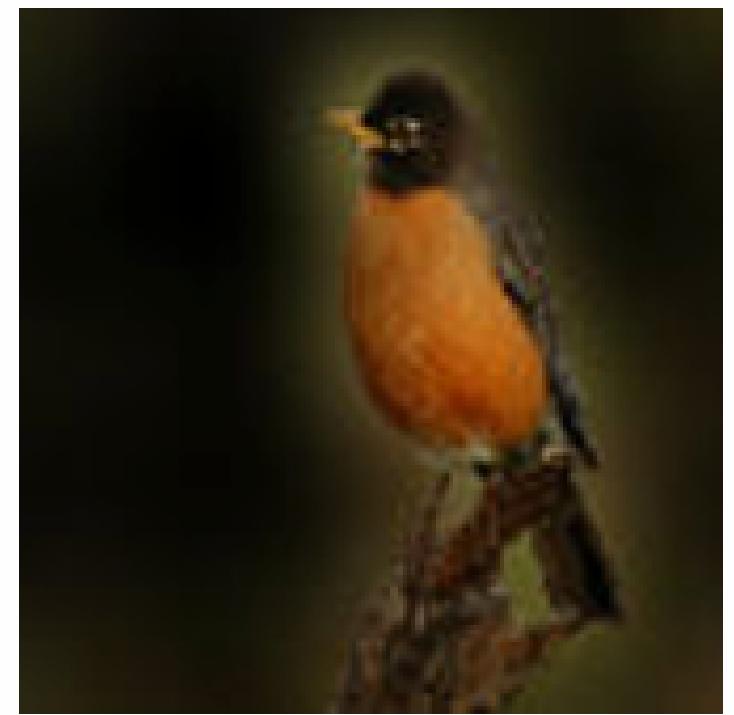
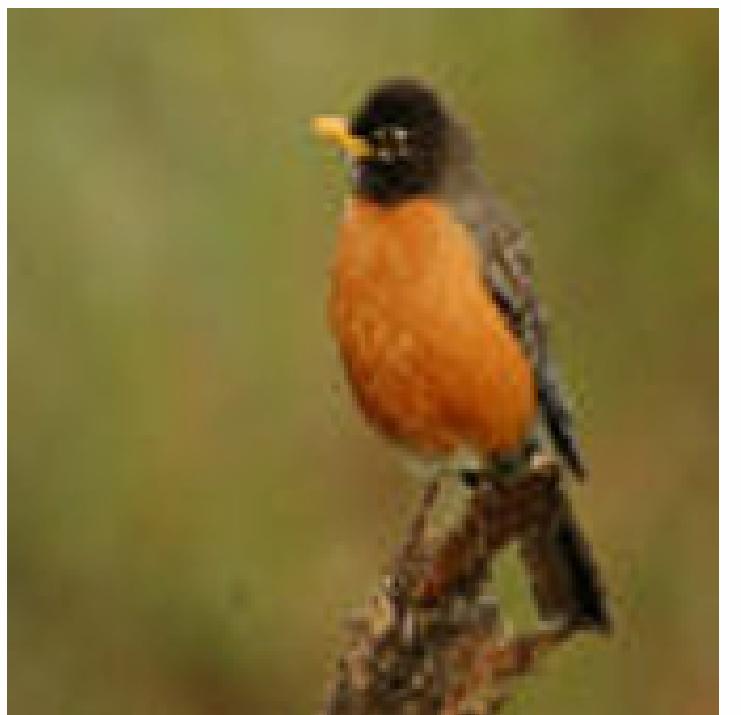
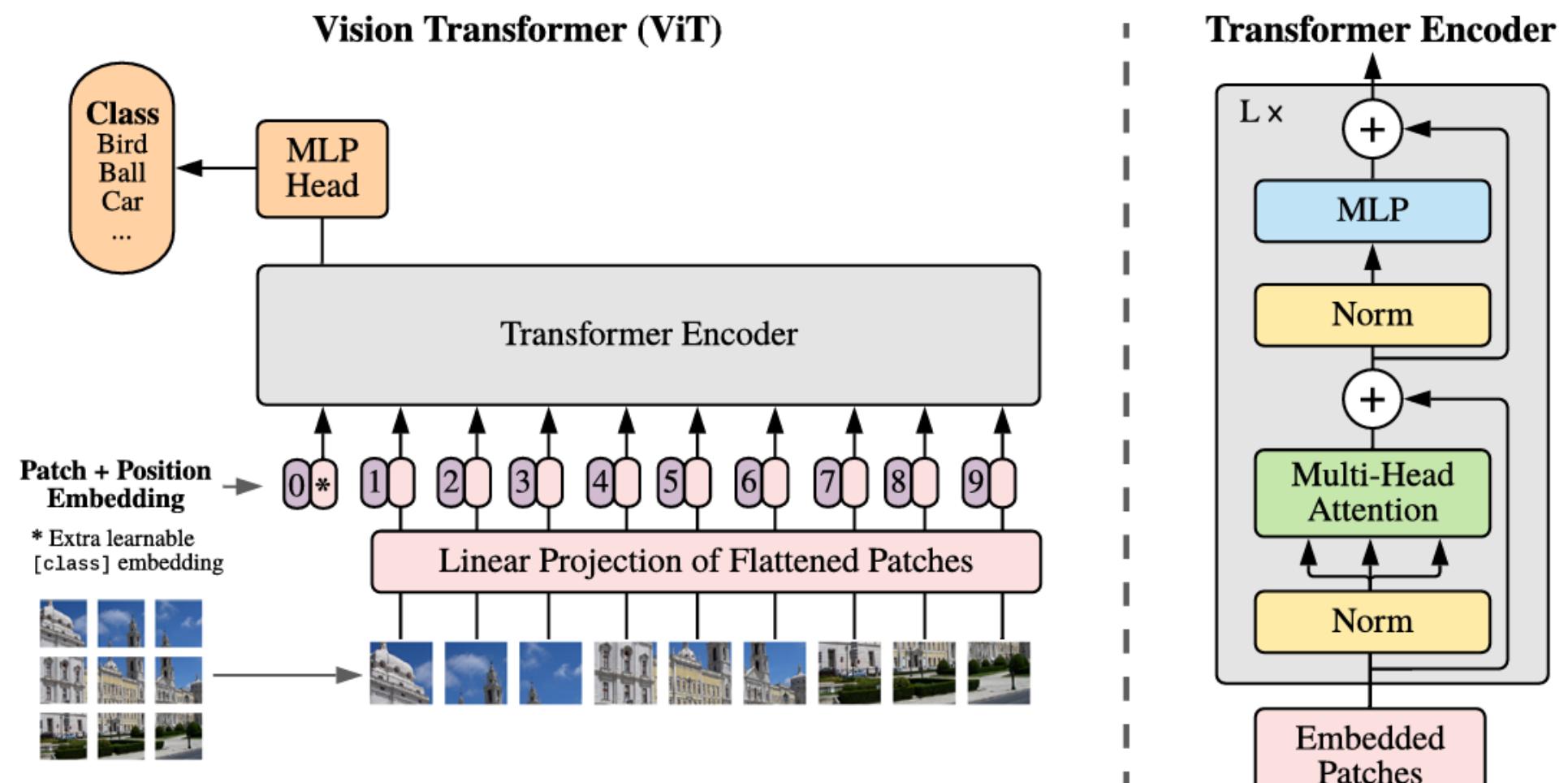
3

Research Paper

Read a research paper An Image is Worth 16x16 Words

VIT ARCHITECTURE AND WORKFLOW

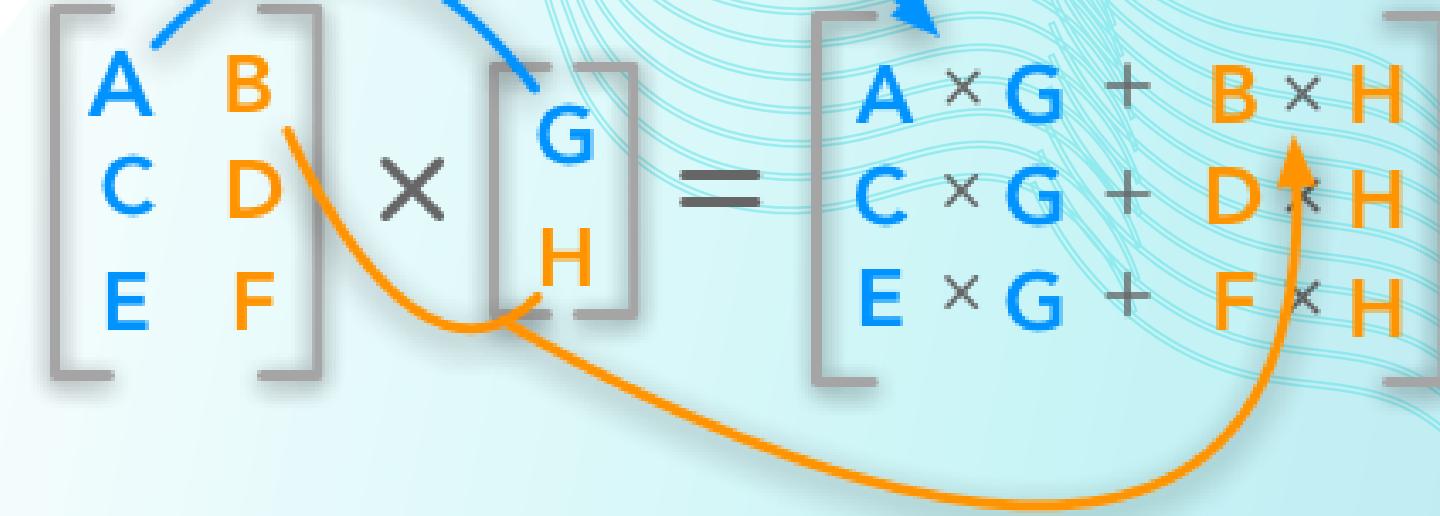
1. Patch Creation : transformation of image into specific shape(224 x 224 x 3) and creation of image patches(16 x 16 x 3).
2. Patch Embedding: linear embedding + position embedding
3. Transformer Encoder: self-attention layer for capturing global context of the image.
4. Classification Head: Soft-max layer gives best predicated image.



PERCENTAGE HITS

In AI/ML Accelerator we will store the pairs which are being repeated so that when they gets repeated we can get it directly from the buffer. And to increase this repetition (percentage hits) we rounded off the weights upto some decimal places. To identify the same pairs, we will follow the logic given beside.

ALGORITHM



- **Step 01**

Convert the whole pair into string and then insert all the strings into a set which will automatically remove same pairs.

- **Step 02**

Now for us (1,2) and (2,1) are same pairs because after multiplication it will give the same result so in order to identify the reverse pair remove a string from the set, reverse it and then again add it into the set.

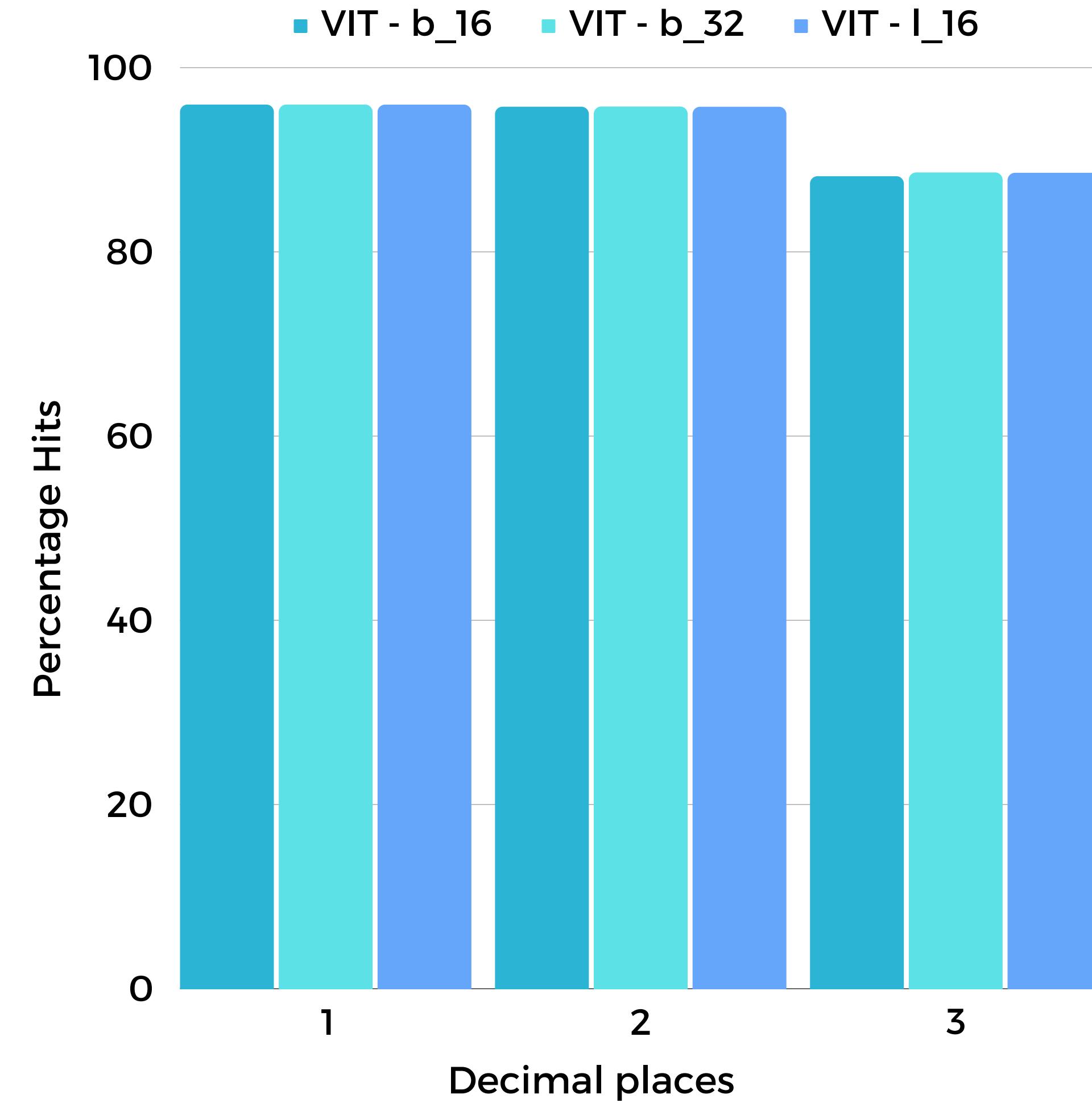
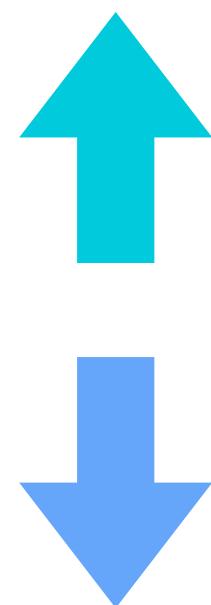
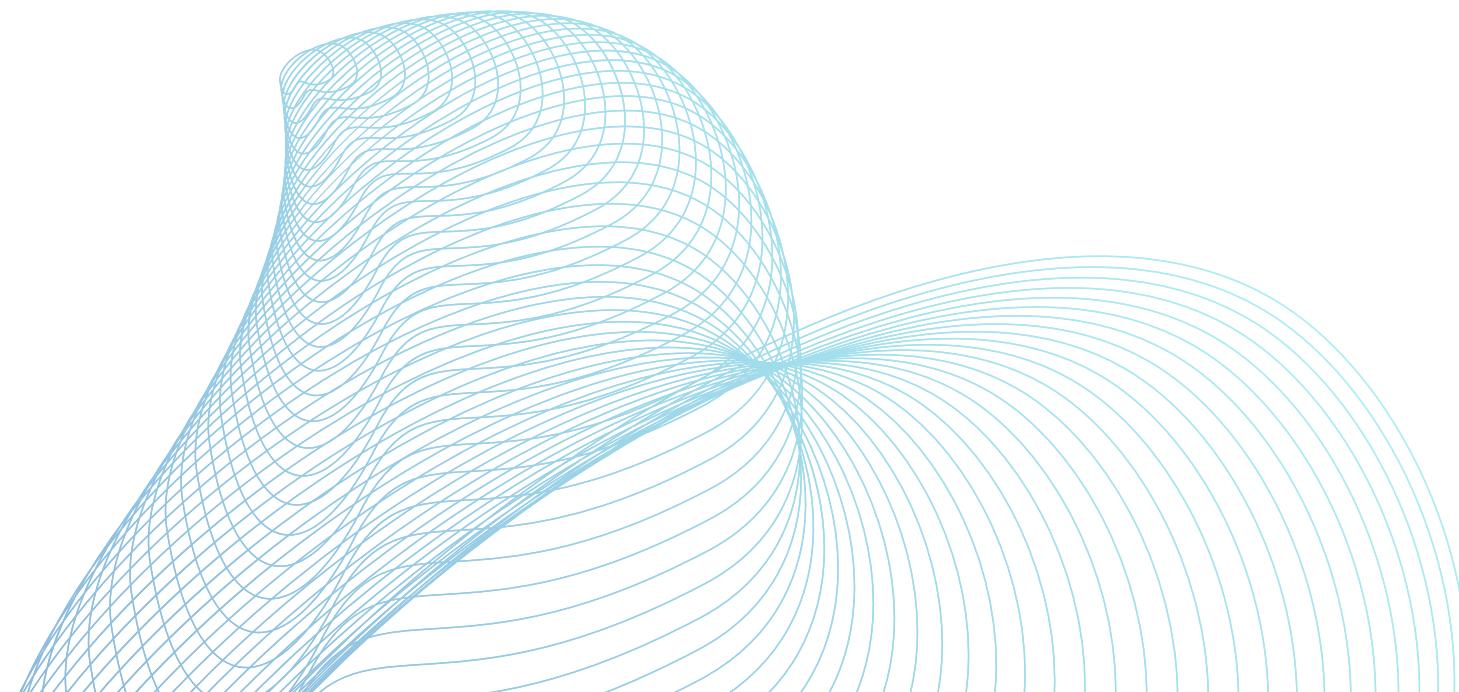
- **Step 03**

Now we will again remove that reversed pair(to avoid the infinite loop) and add into the unique set. By doing this we will reduce the size of previous set either by 1 or 2.

GRAPHS

DECIMAL PLACES

PERCENTAGE HITS

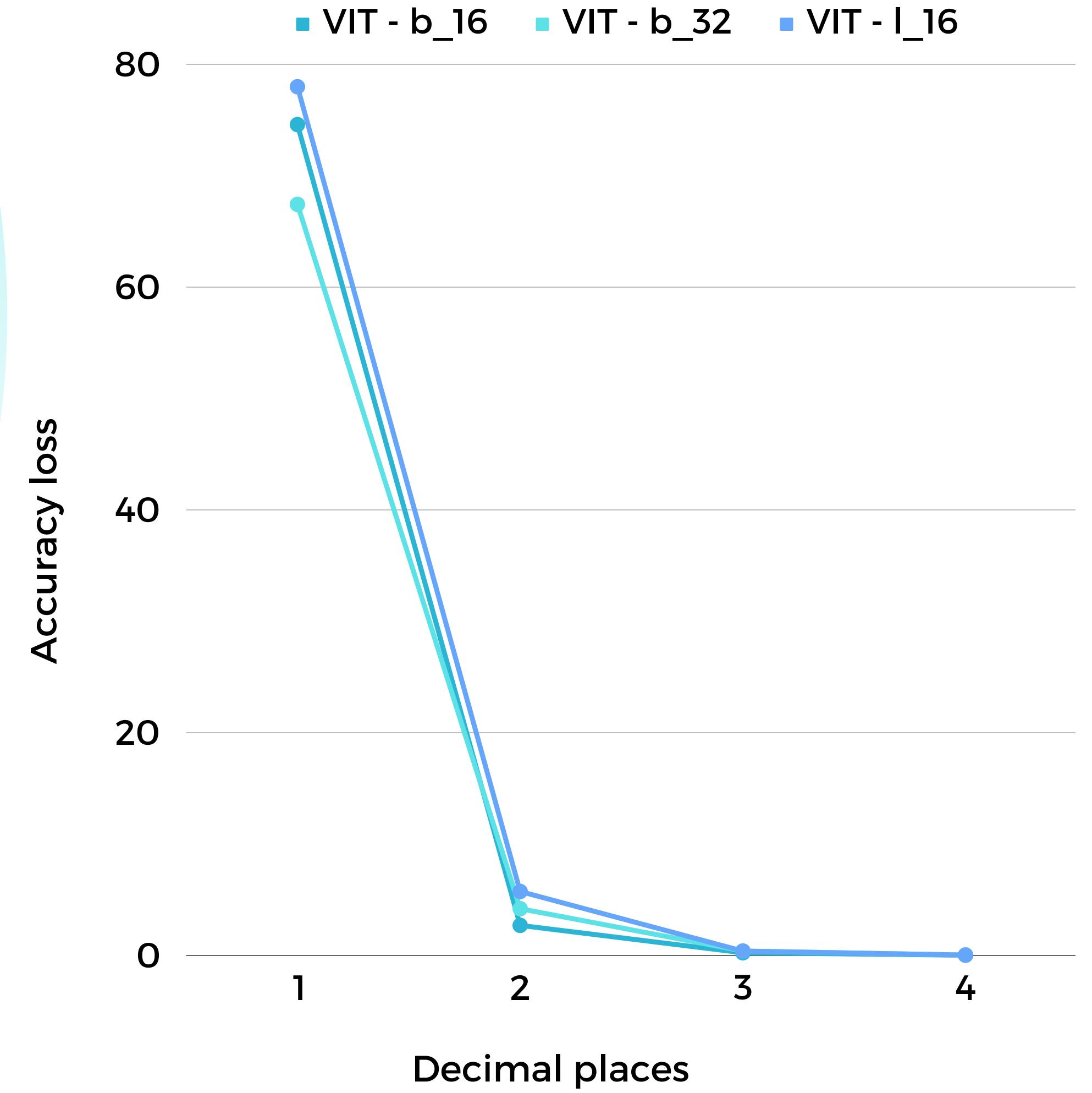
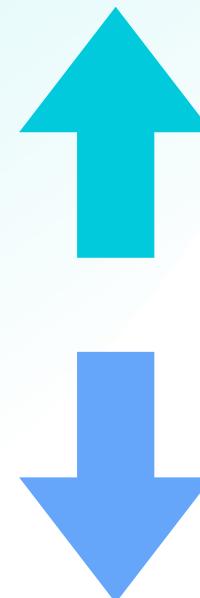


ACCURACY LOSS

Now as we have tempered the weights, this will result in the accuracy loss of model and to calculate this accuracy loss we will compare the accuracy of our model after tempering the weights and compare it with the accuracy of original model.

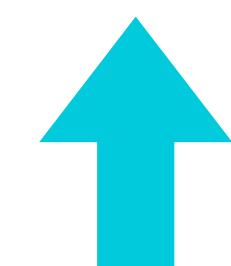
DECIMAL PLACES

ACCURACY LOSS

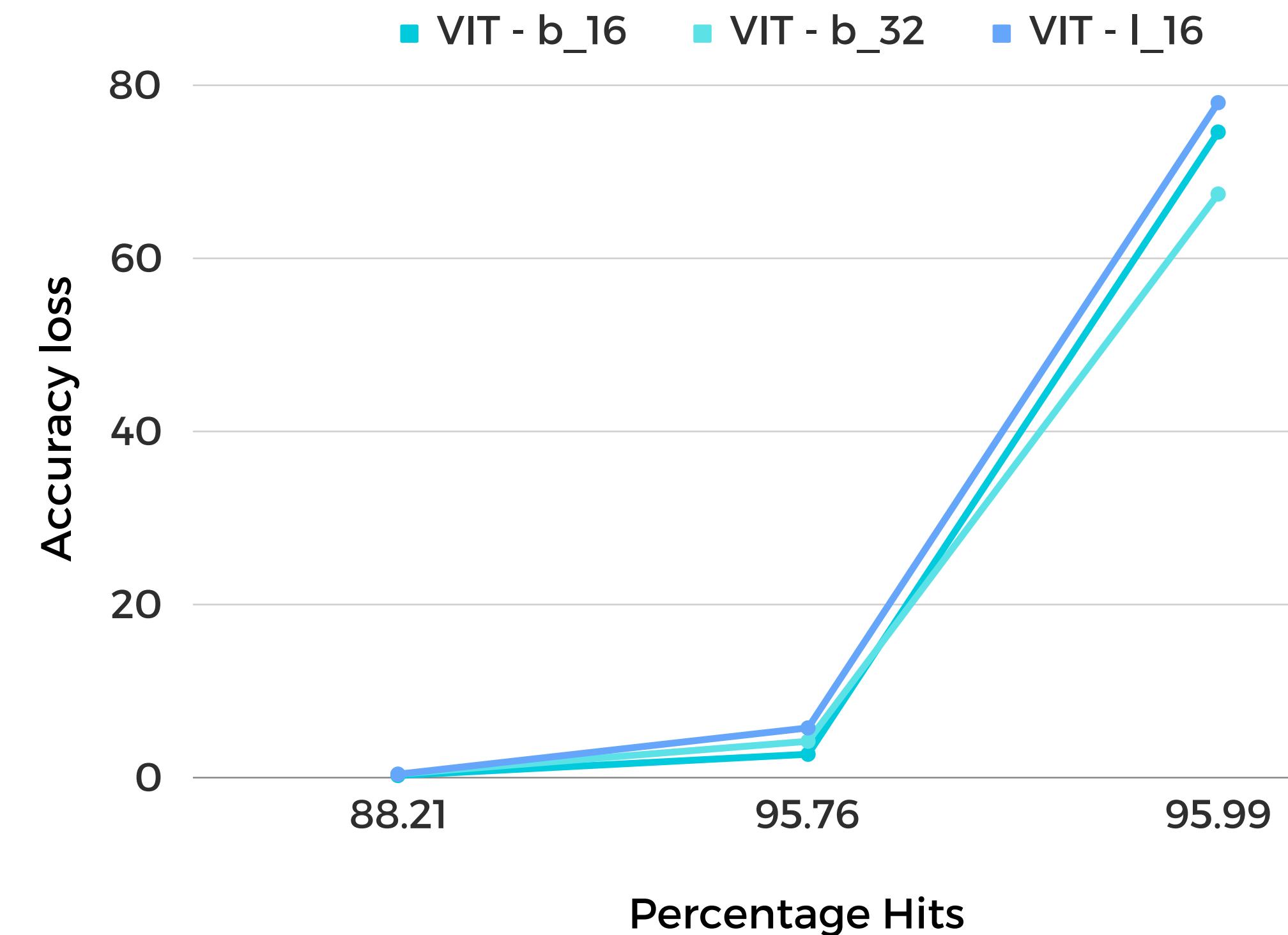
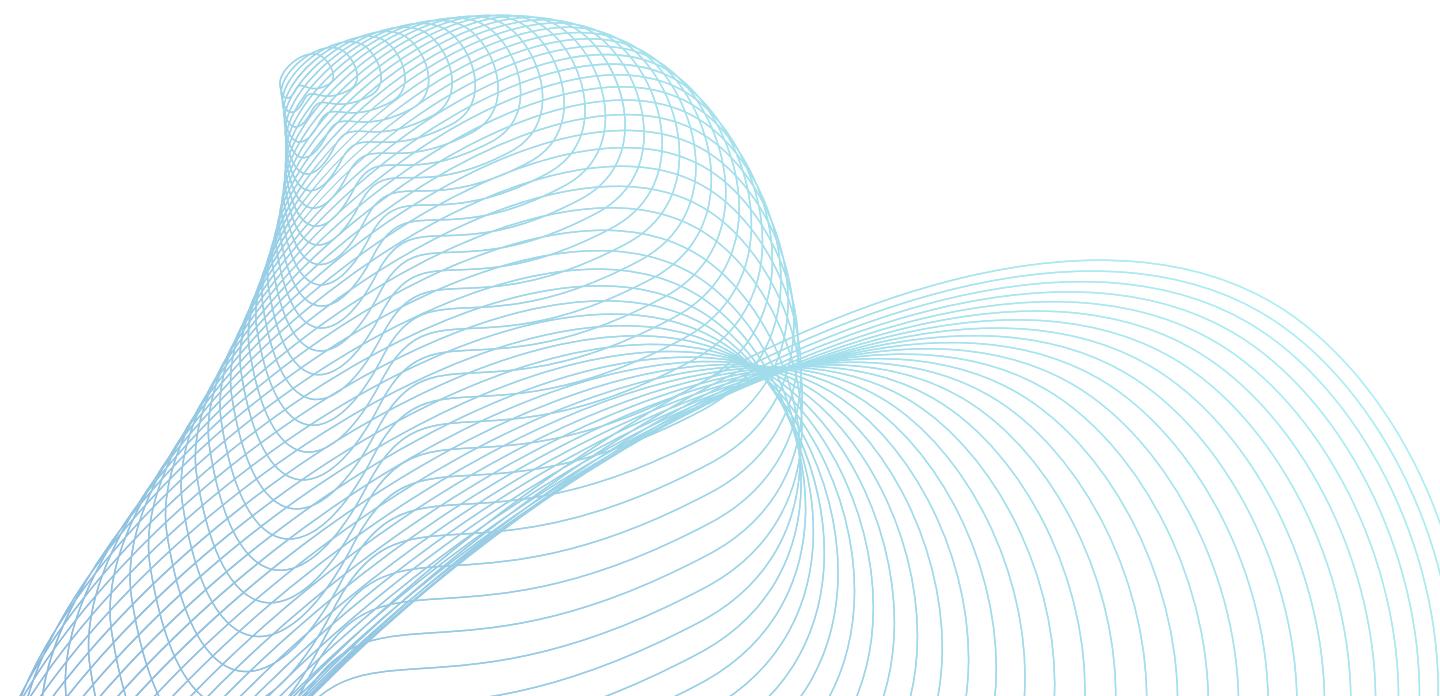


CONCLUSION

PERCENTAGE HITS



ACCURACY LOSS



**THANK YOU
VERY MUCH**

