IJCSE International Journal of Computer Sciences and Engineering Open Access
Technical Paper Vol. 9, Issue.11, November 2021 E-ISSN: 2347-2693

Hybrid ML Recommender System for Visually Similar Product Images

Bagyalakshmi V.[1], Gaurav Sharma[2], Meghna Mahajan[3], Muzzammil Ahmed[4], Kalyan Prakash Baishya[5*], Kuruvilla Abraham[6]

[1]Principal Scientist, Tata Consultancy Services, Chennai, India
[2]Data Science Developer, Tata Consultancy Services, Delhi, India
[3]Domain Consultant, Tata Consultancy Services, Bangalore, India
[4]Data Science Developer, Tata Consultancy Services, Chennai, India
[5]AI&ML Architect, Tata Consultancy Services, Pune, India
[6]Senior Data Scientist, Tata Consultancy Services, Delhi, India

**Corresponding Author: kalyan.baishya@gmail.com*

DOI: https://doi.org/10.26438/ijcse/v9i11.4550 | Available online at: www.ijcseonline.org

Received: 18/Nov/2021, Accepted: 20/Nov/2021, Published: 30/Nov/2021

Abstract- Fashion industry and innovation go hand in hand & technology could not be left far behind when it comes to innovation. Retail fashion is one of the early adopters of artificial intelligence when it comes to product development. AI based applications provides ease of search and shop for products. Either in the form of visual based search or suggesting products from same category with different attributes, retailers are providing every possible easement to customers for better shopping experience. With AI onboard, there is a huge infrastructure cost associated as well. In computer vision (AI), model training requires a good image data with labels & high-capacity platform for starters. Considering these facts, only using transformed feature vector of product images to generate clusters based on feature similarity can reduce the data dependency. Additionally, distance metric can be used to compute the feature distances & retrieval of top-k similar images by reverse indexing of image features to their corresponding images.

Keywords- CNN (Convolution Neural Network), AI (Artificial Intelligence), Image Processing, Clustering, Feature Extraction, Unsupervised image-based recommender System.

I. INTRODUCTION

Application of artificial intelligence has a profound imprint on the customer facing applications in businesses. Images and videos form a very pertinent input using which artificial intelligence could be imbibed in applications. This state of artificial intelligence often relies on developing complex neural networks producing human like accuracies (at the expense of a lot computational power). The source of inspiration for this complex architecture is said to have derived from the human brain, considered to have multi-dimensional processing capability.

Human vision system is highly capable of detecting patterns in a product/object in few shots and then can spot visually similar products from a gallery (or) catalogue with higher accuracy.

Computer vision aims to achieve this capability through set of algorithms. Computer vision is an area of artificial intelligence which enables computers/systems to process image or video data and derive meaningful information from them.

Some essential computer vision terminologies used in the paper are briefed below:

Supervised learning
It is a machine learning approach in which an algorithm is trained on labeled data wherein the function learns to map the processed input data to the output label(s).

CNN (convolutional neural network)
A general cnn structure consists of an **input layer** through which a product/object image is fed to the model as image matrix. The image matrix is passed onto series of **convolutional layers** which then apply matrix convolution operations and obtain feature-maps (fms).

These fms essentially contain pixel level information on parts of the product/object to be focused on. These fms are then passed onto **pooling layers** which filters out not-so-useful features from them. These convolution-pooling operations are repeated as many times as needed.

Eventually pooling layer output is fed to **fully connected** (fc) layer. This fc layer flattens the pooling layer output and then maps the feature-vector (fv) to the ground-truth (gt) values using **dense layer**.

In cnn model training stage, values of kernel filters of all the layers (as mentioned above) are adjusted/modified to retrieve the desired feature information from each product image and thus the feature-vector (fv) obtained on fc layer is representative of that desired product feature information. So, we will be removing final dense-layer(s) to obtain the desired feature-vectors.

Deep convolutional neural networks (cnns) have been widely used in variety of computer vision-based business problems especially in retail domain. One of them is product deep feature extraction. Numerous cnn based models have been developed for the same and some of them have been established as state-of-the-art (sota). Those sota models include inception series, xception, efficient net series etc., efficient net series offers the flexibility to choose a less-complex yet efficient cnn structure to perform the desired deep-feature extraction. Here, model complexity is measured in terms of number of trainable parameters. And so, the efficient net model has been chosen for the purpose.

Pre-trained models

Pre-trained models are models solved by someone else to solve a similar kind of problem, instead of building a model from scratch we can used the already trained model use it as a starting point and use it for customizing the same.

Attribute's extraction

It is a process of deriving meaningful features from existing dataset which may be informative or useful for facilitating the learning and generalizations steps and will help in carrying out more relevant output useful for the analysis.

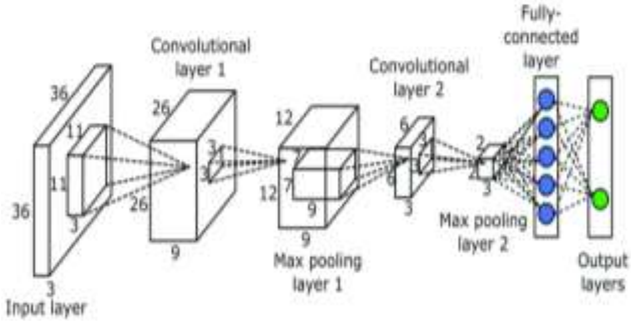


Figure 1: General CNN Architecture [1].

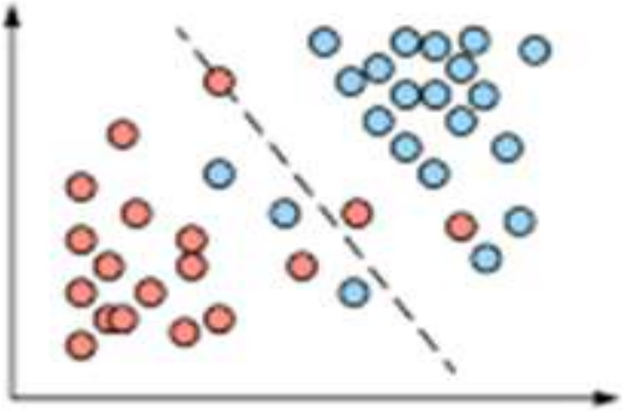


Figure 2: Supervised Learning

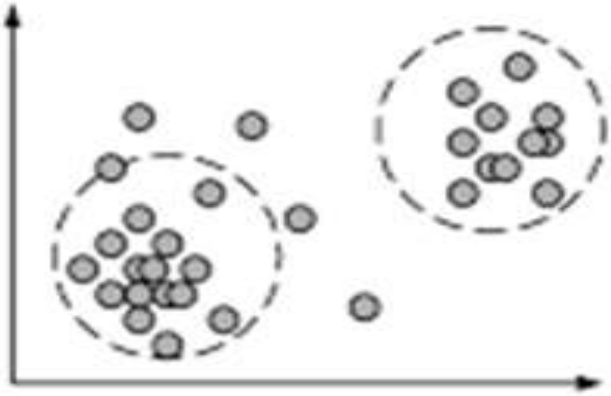


Figure 3: Unsupervised Learning

Unsupervised learning

It is a machine learning approach in which an algorithm is trained on unlabeled data wherein the hidden patterns in the input data are discovered which is subsequently grouped/clustered. Unsupervised way of analyzing has also been proven to be an effective method to be used in recommender system in grouping the feature vector information in product catalogs. With the development of recommender systems and unsupervised techniques, many clustering algorithms have been used to solve vision-based search and similarity-based problems.[2]

K-means clustering

To make the process more efficient and accommodating towards unlabeled data in the database, k-means clustering is used. K-means is an unsupervised learning algorithm which is used to find clusters in the data without having prior understanding of the data points' properties. It uncovers similarities based on which grouping/clusters are created. K represents the number of clusters or number of cluster centers. Feature similarity is used as a measure to classify the data into clusters with similar data points belonging to a cluster and dissimilar belonging to others.

The above Figure 2 shows a classification task (supervised approach) for samples between two random variables, where class labels are shown in scatter plots separated within a linear line or boundary separating two colours. Therefore, in supervised learning, class labels used to build the classification models are known in advance.

As shown in Figure 3, the classes (or circles) shown above are all same colour or not known in advance and can be inferred as an unstructured dataset. In contrast, unsupervised learning task deal with unlabelled dataset instances as classes are not known in advanced to the model.

Visualizing high dimensional clustered data

Human eyes perceive, at best, in four dimensions. Visualizing dimensions more than four often requires a reduction in the number of dimensions from many to the two (or three) dimensions that best allows for visual interpretation. Typically, this is done through dimension reducing approaches like principal components analysis (pca) but of late, t-SNE has become a preferred technique for embedding highly dimensional data in to lower dimensional space while preserving local structure. And this can be used in categorizing images based on their features.[3]

II. PROPOSED METHODOLOGY

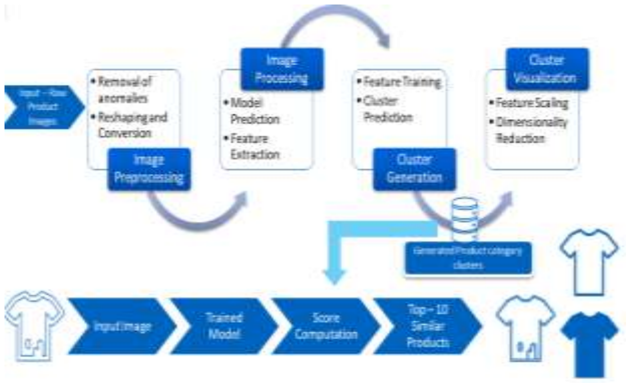


Figure 4: Approach Flowchart

1) *Data Preparation*:
In Machine Learning gathering & preparing data for a model is very crucial as well as a sensitive task. According to various objectives, the need & quality of data varies. This phase includes image accumulation & preprocessing.

a) *Data Gathering*:
The Image data has been prepared by scrapping images from few women's fashion retailers. The data consist of 14 major categories such as top-wear, jeans, swimwear, watches, bracelet, suits, coats, skirts, etc.

b) *Data Cleaning*:
This step includes searching for those images which are of bad quality, poor resolution, etc.

c) *Monochromatic–Background Consistency*:
All the images have been analyzed for presence of multi-objects/multi-color variable inconsistent background so that these irrelevant images can be removed.

d) *Data Reshaping*:
As the last step of preprocessing, the images have been reshaped to the shape desired by Feature Extractor i.e., (224,224,3).

2) *Feature Prediction & Extraction*:
For the prediction of image level features, Efficient- Net Model has been used as a feature extractor. In order to make classifier as a feature extractor, top layer of the model is removed and SoftMax as an activation layer is appended.

a) *Feature Prediction*:
After preprocessing the images, the data has been passed to the customized feature extractor, and after processing of data, model predicted features as an output. The image array, after passing to the feature extractor transformed into feature vector.

b) *Feature Extraction*:
The output shape of feature extractor is complicated to visualize and use it for model training. This part includes reshaping of feature extractor output and make it feasible for unsupervised model training.
After this phase, the final output has more than 80k features per image.

3) *Features Training & Cluster Prediction*:
Here, images & its features are the only input data. This is considered to be non-labeled data because images have no corresponding target value associated with it. This is a typical unsupervised setting of machine learning.
This technique is used to train models on feature extractor output.

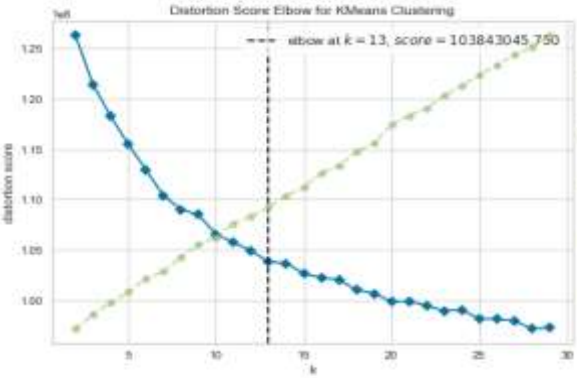


Figure 5: k (No. of Clusters) Vs Distortion score Elbow Plot

a) *K-means Model Training*:
For this step, k-means model has been initialized with k= 2 and goes up to 30. For making a consensus over value of 'k', K-Elbow Visualizer has been used. Fig 3 shows results corresponding to different values of k & generating distortion score helped in understanding the model behavior for various cluster counts on images features.

After analyzing Fig 5, it can be said that due to the high similarity among image features of all images, elbow plot is still not able to provide crystal clear picture of the value of k which needs to be considered. But since this is completely unsupervised way of model training, even if we get some cue for k, it should be good enough.

While the elbow plot is not able to accurately pinpoint the appropriate values of k to be selected, it does provide us with some very good directions. From the plot, it could be observed that the elbow bend occurs somewhere between values of k lying between 13 to 15. And when we try to compare this observation with the total number of image categories we originally had, we could grasp that the predicted k value (from the graph) lies closer to the actual range which is 14

b) *Cluster Generation*:
After training two instances of k-means model for k=13 & k=14, the clusters have been generated for both models. Image Features have been allocated to the corresponding cluster based on the feature's similarity. For visual valida-

tion of each cluster under both models, images need to be segregated in their corresponding clusters.
After using reverse indexing technique, the images have been segregated to their respective predicted clusters for the whole data.

4) *Data Visualization*:
In an unsupervised setting, determining model's accuracy is a strenuous exercise and not fairly straight-forward. However, there are very good visualization techniques which we can rely upon and at the same time handle high dimensional data exceptionally well. We are going to fall back to one of the most preferred visualization techniques known as t-SNE plots to validate our results. Visual based techniques are endorsed over metric based as they lend credibility of ease of explanation and proof.

a) Dimensionality Reduction:
This step includes depletion of Image features from 87k to 2 data points for each image for visualizing the data on 2-d chart. But removing features is not a straightforward task, it may affect critical & decisive features as well.

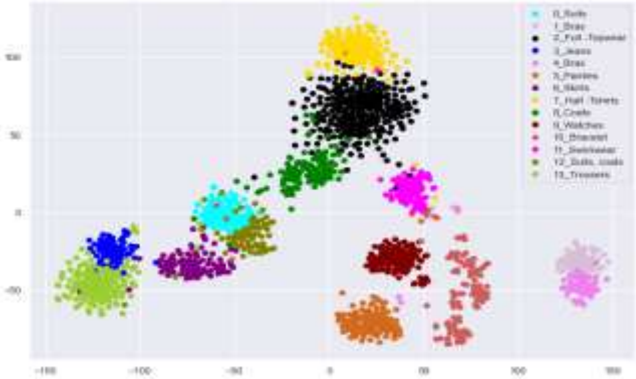


Figure 6: t-SNE Plot of Image Features at Cluster level distribution

This process should be gradual and considerate. To fulfill this requirement, t-SNE (t-distributed Stochastic Neighbor Embedding) approach has been used.

b) *t-SNE Initialization & Training*:
For better visualization of high dimensional data, t-SNE Model has been trained on the feature extractor outputs. The model has been initialized with principal component analysis (PCA) parameters to begin with its first iteration. t-SNE outputs has been assigned to the cluster values for the representation of features along with the respective cluster number.

In Figure 6, we can clearly observe the inter-cluster diversity, even though there are some overlapping of data points which points to ultra-high similarity among few products.

5) Fetching top – k Visually Similar Products:
Now, after generating clusters for the whole data, trained model can be used for recommending top – k products that are visually similar to the test image which have been passed to the model.

a) *Image – Reshaping*:
In order to extract features from image, it has to be passed into the feature extractor. Passing a test image to feature extractor requires a specific dimension of image i.e., (224,224,3). This step includes reshaping the test image and make it suitable for processing.

b) *Feature Extraction*:
After reshaping the test image, the image array has been passed to the feature extractor and transformed into feature vector.

c) *Cluster Prediction*:
The feature vector generated from feature extractor, now passed to the trained k-means model. For the test image, feature vector corresponding cluster has been predicted by trained model.

d) *Distance – Matrix Generation*:
For the feature vector of test image, distance matrix has been computed for each image that belongs to the same cluster as of test image.

e) *Fetching Products*:
From the sorted distance matrix, top k minimal distance feature vector indexes have been fetched from the matrices. For recommendation of images, reverse indexing is used to retrieve the corresponding images. As a final output of the model, Figure 5 is a sample that can be consider as an objective to achieve.

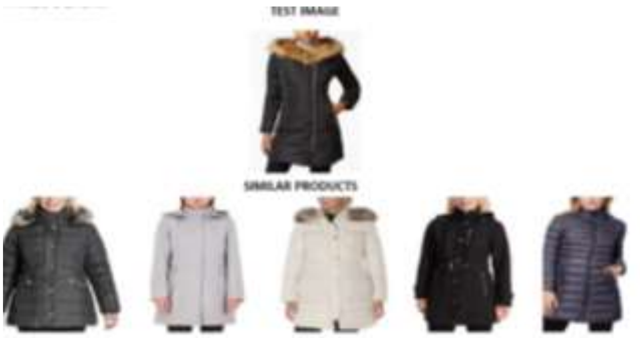
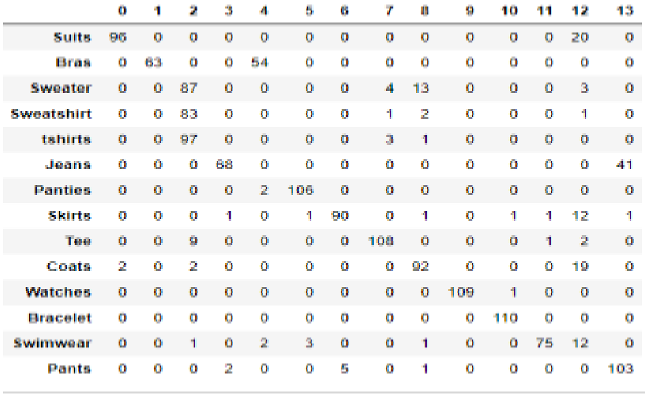


Figure 7: Sample Input – Output of the Model

III. CONCLUSION



	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Suits	96	0	0	0	0	0	0	0	0	0	0	0	20	0
Bras	0	63	0	0	54	0	0	0	0	0	0	0	0	0
Sweater	0	0	87	0	0	0	0	4	13	0	0	0	3	0
Sweatshirt	0	0	83	0	0	0	0	1	2	0	0	0	1	0
tshirts	0	0	97	0	0	0	0	3	1	0	0	0	0	0
Jeans	0	0	0	68	0	0	0	0	0	0	0	0	0	41
Panties	0	0	0	0	2	106	0	0	0	0	0	0	0	0
Skirts	0	0	0	1	0	1	90	0	1	0	1	1	12	1
Tee	0	0	9	0	0	0	0	108	0	0	0	1	2	0
Coats	2	0	2	0	0	0	0	0	92	0	0	0	19	0
Watches	0	0	0	0	0	0	0	0	0	109	1	0	0	0
Bracelet	0	0	0	0	0	0	0	0	0	0	110	0	0	0
Swimwear	0	0	1	0	2	3	0	0	1	0	0	75	12	0
Pants	0	0	0	2	0	0	5	0	1	0	0	0	0	103

Figure 8: Instance Level - Output Validation

Figure 7 represents the feature level cluster visualization of complete data. By seeing the product images, it can be said that the difference among product categories is quite visible. This difference is proportional to the super category of products. As the product categorization goes down with the specialization level, the similarity gets increased and so the model find difficulty finding the variation among image features. But when categorization goes with generalization level/super category level the inter class variation get increased and model find it easy to determine the feature level difference.

In figure 8, 14 clusters with their corresponding product category instances have been evaluated. The table represents clusters (0 -13) &14 Product category (Bras, Suits…. Pants). Each product instance has been validated against its product category & the cluster it belongs. For categories like Tee, Bras, Panties, Watches & Bracelet, the model almost performs more than 90% since it clusters most of the products in same cluster. For Other categories model has shown the performance of 75-80%.

This very good performance has been achieved with no additional cost of high-end resources. So, from outputs it can be stated that Unsupervised learning with Pretrained Models can be a useful approach to achieve considerable amount of performance when there is slight lack of balance in data quality & computation resources.

The experimental outcome of this activity is commendable. With less resources & unavailability of the labelled data, this approach becomes a good case of utilization of pretrained networks along with merging with other machine techniques learning approaches for achieving a low-cost recommender system. Since only Image features has been trained here with no target class values, this model additionally, can be scaled for more diverse class images such as sarees etc.

The work that has been accomplished strives to bring value for the retailers while displaying products in their websites with minimum cycle time. It also lends greater flexibility to the buyers while making a purchase decision which in turn enhances the brand value. As it has been, understanding customer buying behavior helps a long way to position products.

In order to catch up with the ever-growing pace of the fashion industry, it becomes imperative for a retailer to bring the latest offering to its customer and the same point maintain the level of personalization as well.

It is often seen that the retail industry is a great source of wide variety of data generation, and it make perfect sense to utilize them leveraging artificial intelligence (read machine learning). Retailers aspire to create concordant experience for the buyers with the amalgamation of their product lines and buyers' preferences. This in turn helps them monetize data efficiently.

As time progresses, more retailers are expected to adopt artificial intelligence-based offering to win over customers and increase revenues. Ease of decision-making while purchasing products has an everlasting effect on the customers. This work presents a win-win situation for both customers and retailers. It also provides an opportunity for a retailer to understand its shortcoming compared to the competition.

IV. FUTURE WORK

As of now, current experiments have been made to target the fashion wear in the retail segment. However, this could also be replicated to different other areas of retail industry like home & furniture, accessories, footwear with proper modification in our existing work. Hence, this work has the capability to scale up and cater to the entire segments of retail business.

REFERENCES

[1]. B. P. Amiruddin and R. E. Abdul Kadir, *"CNN Architectures Performance Evaluation for Image Classification of Mosquito in Indonesia"*, 2020 International Seminar on Intelligent Technology and Its Applications (ISITIA), **2020**, pp. 223-227, doi: 10.1109/ISITIA49792.2020.9163732.

[2]. Aayush Kumar Singh, Abhishek Kumar and Kuldeep. *"Image to Image Search using K-means Clustering".* International Journal of Computer Applications (0975 – 8887) Volume **182** – No. 46, pp. - March **2019**

[3]. D. M. Chan, R. Rao, F. Huang and J. F. Canny, *"T-SNE-CUDA: GPU-Accelerated T-SNE and its Applications to Modern Data",* 2018 30th International Symposium on Computer Architecture and High-Performance Computing (SBAC-PAD), **2018**, pp. 330-338, doi: 10.1109/CAHPC.2018.8645912.

[4]. Yash Baid, Avinash Dhole, *"Food Image Classification Using Deep Learning Techniques",* International Journal of Computer Sciences and Engineering, Vol.**9**, Issue.**7**, pp.11-15, **2021**

[5]. Astha Pathak, Avinash Dhole, *"Image classification Method in detecting Lungs Cancer using CT images: A Review",* International Journal of Computer Sciences and Engineering, Vol.**9**, Issue.**5**, pp.37-42, **2021**.

AUTHORS PROFILE

Mr. Gaurav Sharma pursued Bachelors of Engineering in Computer Science & Engineering from RGTU Bhopal in year 2014. He also has a M. Tech from Computer science allied to his U.G. He is currently employed as Machine learning developer at Tata Consultancy Services. His main work comprises of Machine learning / Deep learning-based prototypes (capability) development in the internal R&D project. He has been in the domain of Data Science since June, 2020.

Mrs. Bagyalakshmi V. has been working as a Principal Scientist at Tata Consultancy Services, Chennai for over 15 years. She has completed her B. Tech degree in Electrical & Electronics Engineering from Government College of Technology, Anna University, Coimbatore. She also has a Post Graduate from Robotics allied to her UG.
She is a pronounced Data Science enthusiast with many research activities & publications in the field of Artificial Intelligence to her credit.

Mr. Kalyan P. Baishya has been working as Machine Learning & Artificial Intelligence professional with over 10 years of experience in Tata Consultancy Services. He has provided consultation and implemented Artificial Intelligence based solutions across Fortune 500 companies mostly in Banking, Oil & Gas and Retail sector. He has pursued Master of Business Administration from Amity Business School, Noida in Marketing & Operations after completing Bachelors of Engineering in the field of Mechanical from Jorhat Engineering College, Jorhat, Assam.

Ms. Meghna Mahajan has pursued Bachelors of Engineering in Electrical engineering in the year 2014 and completed her MBA in Marketing and HR. She is associated with Tata Consultancy Services since 2017. She holds 3 years of experience as SAP-SD functional consultant Currently she is the domain consultant for Retail sector.

Mr. Muzzammil Ahmed K. pursued Bachelors of Engineering in the field of Instrumentation and Control from PSG College of Technology, Coimbatore, Tamil Nadu in year 2018. He is currently employed as Machine learning developer at Tata Consultancy Services. His main work comprises of Machine learning / Deep learning-based prototype (capability) development in the internal R&D project. He has been in the domain of Data Science since Jan, 2019.

Mr. Kuruvilla Abraham pursued Bachelors of Engineering in Electronics & Communication Engineering from Rajasthan technical University, Kota in year 2011.He also has a M. Tech in Computer science from BITS, Pilani, Rajasthan. He holds 8 plus years of experience in Analytics in areas of Machine learning, Deep learning in Computer Vision, Embedded robotics, Natural language processing, Business Intelligence for multiple industries across various geographies.

Primary responsibility for complex business problem from data and programming, generating new ideas and innovations. He is currently employed as Senior Data Scientist at Tata Consultancy Services. His main work comprises of Machine learning / Deep learning-based prototypes (capability) development in the internal R&D project. He has been in the domain of Data Science since 2015.