A Report on

INPLANT TRAINING

AT

Jio Platforms Limited (JPL)

Reliance Corporate Park, Thane - Navi Mumbai - 400701

Submitted By

Vansh Shah

SAP ID: 70321019094

Under The Guidance Of

Prashasti Kaniker

Shri Vile Parle Kelvani Mandal's

Mukesh Patel School of Technology & Management Engineering

Department of Computer Engineering

Vile Parle (w), Mumbai- 400056

Inplant Training Period

2nd January 2023 to 28th April 2023

NO OBJECTION CERTIFICATE

This is to certify that Vansh Shah (SAP ID. 70321019094) student of Mukesh Patel School of Technology Management & Engineering, has successfully completed his/ ~~her~~ Semester **(VIII/ ~~XI~~)** Inplant Training from 2nd January 2023 to 28th April 2023 , at Jio Platforms Limited. (JPL) and has completed the tasks given to him during the Inplant Training. The Report is submitted in partial fulfillment of the requirement for B.Tech Integrated Computer Engineering. He/~~She~~ have been allowed to include the relevant information from the Company for which we have no objection.

Signature & Stamp of The Industry

Name: Aditya Chauhan
Email ID: aditya3.chauhan@ril.com
Contact: +91 9326222811
Date: 04-05-2023

CERTIFICATE OF SUBMISSION

This is to certify that the report entitled 'Inplant Training at Jio Platforms Limited (JPL)' is a bonafide work of Vansh Shah (SAP ID. 70321019094) and submitted in semester (**VIII/ ~~XI~~**) for the partial fulfillment of B.Tech Integrated in Computer Engineering at, Mukesh Patel School of Technology Management & Engineering Mumbai during the academic year 2022-2023.

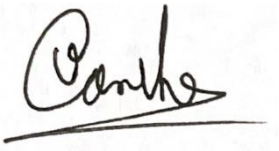
_____ _____ _____

Signature of the Faculty Mentor **Signature of the HoD** **Signature of the Dean**

DECLARATION

I, Vansh Shah, SAP ID. 70321019094 a student of Semester (**VIII/ ~~XI~~**) of B.Tech Integrated Program in Computer engineering humbly submit that I have completed from time to time the Inplant Training work as described in this report by my own skills and studied from 2nd January 2023 to 28th April 2023. I have not copied the report or it's from any appreciable part from any other literature in contravention of academics ethics. I declare that this written submission represents my ideas in my own words. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented of fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be the cause of disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Date: 28th April 2023



Place: Navi Mumbai

Name & Signature of the Student

Name: Vansh Shah
Email ID: vansh.shah94@nmims.edu.in
Contact: +91 9967316669

- **Cover Page**
- **No Objection Certificate**
- **Certificate of Submission**
- **Declaration**

Chapter 1 – Introduction

1.1 About the company:

Jio Platforms Limited is a leading Indian telecommunications company that provides a wide range of digital products and services. Established in 2010 by Mukesh Ambani, the chairman of Reliance Industries, Jio Platforms is a subsidiary of one of India's largest private-sector conglomerates.

Jio Platforms Limited provides solutions for customers (B2C) and enterprise (B2B). Services include an end-to-end 5G solution consisting of 5G Radio, a complete 5G Core Network, AI/ML ATOM platform for 4G/5G, MANO for cloud CNF orchestration, ACI for cloud infrastructure deployment as well as the Cloud-native OSS Platforms.

In addition to these core platforms, JPL has also created own cloud-native probing solutions for radio and core networks, simplifying network debugging and not requiring any systems integration with probe providers. To enable a SAAS-based model, JPL has also developed a sophisticated array of BSS solutions.

Fig 1. Company Logo
Jio Platforms Limited

1.2 Background of Industry:

Over the past decade, the telecommunications industry in India has experienced explosive growth, with Jio Platforms leading the way. Through its innovative products and services, Jio Platforms has been able to capitalize on India's vast telecommunications market, becoming one of the country's premier digital solutions providers.

1.3 Products & Services:

At the forefront of Jio Platforms' product offerings is Jio, a wireless broadband service that provides high-speed internet, voice and video calling, and digital content. In addition to Jio, the company has launched several digital apps and services, such as

- JioTV
- JioCinema,
- JioSaavn,
- JioMart.

These apps offer customers digital content, including movies, TV shows, and music.

Jio Platforms' focus on developing a digital ecosystem positions it to meet the ever-evolving needs of its customers across India, and the company remains committed to expanding its product and service offerings.

Overall, Jio Platforms Limited's strong presence in the Indian telecommunications industry and innovative digital solutions continue to set it apart from competitors. As the company continues to grow and evolve, it remains a crucial player in the digital revolution in India and beyond.

Fig 2. Current Suite of
Products & Services Offered

1.4 About the Department

The department "Special Initiatives AI/ML" focusses on developing cutting edge services and products for both B2B and B2C markets. By incorporating the latest technologies such as AI, ML and DL new products and services are created that can transform the lives of people and business on India. The department is staffed with experienced professionals who have valuable knowledge, skills and expertise in AI and ML technologies.

One of the key areas of research and development is the application of Computer Vision assisted workflows and products. Computer Vision is a field of AI that allows a computer to percept its environment and derive valuable details and data using specialized sensors. Computer Vision also deals with image analysis and processing. The department is working and actively developing new Computer Vision based technologies such as object tracking that can be incorporated into various fields such as retail, autonomous security systems, and infrastructure monitoring.

In addition to these areas, the department is also actively experimenting and exploring with other emerging technologies such as Generative Adversarial Networks (GANs), Reinforcement Learning and Deep Learning. This allows the company to be one of the front runners in the field of AI and ML research and development.

Concluding, "Special Initiatives AI/ML" is a crucial part of the company's strategy to continue its journey as one of the biggest names in the digital space of the Indian Market as well as to become a pioneer and adopter of the new AI/ML capabilities and technology. With a strong focus on the new and upcoming technology, the company aims at producing innovative and market disrupting products that will transform the lives of various people and business sectors in India.

Chapter 2 – Technical Stack and Hardware

2.1 Technologies

The Special Initiatives AI/ML department at Jio Platforms uses a range of advanced technologies, equipment, and hardware to research, develop, and implement its AI and ML-based products and services. The key technologies used include:

2.1.1 Artificial Intelligence (AI)

AI refers to the technology which makes a computer program capable of emulating human behaviors such as

- Reasoning
- Adaptation
- Learning

Computer programs can essentially think and exhibit human-like behavior for their specific tasks. The department uses AI and its subfields as their backbone to build smart products.

2.1.2 Machine Learning (ML)

Machine learning is a subset of AI that deals in developing algorithms which allow a program to:

- Understand provided data.
- Learn its underlying trends.
- Make predictions based on provided data.

The department develops such programs by using machine learning techniques such as supervised, unsupervised, and semi-supervised learning, reinforcement learning etc. to develop robust self-adapting programs.

2.1.3 Computer Vision (CV)

Computer Vision is a field of AI that allows a computer to percept its environment and derive valuable details and data using specialized sensors. Computer Vision also deals with image analysis and processing. It can be incorporated into:

- Autonomous monitoring systems
- Automatic tracking systems
- Image data analysis
- Image Classification and Segregation

It is one of the key technologies used by the department for creating new products and services that rely on visual sensor data and feeds.

2.1.4 Deep Learning (DL)

Deep learning is a subset of machine learning that uses multiple layers to progressively extract higher-level features from raw input. It is based on artificial neural networks. Learning and can be:

- Supervised
- Semi-supervised
- Unsupervised.

Deep learning also forms the backbone of many features added to the products created by the department.

2.1.5 GPU Acceleration Libraries

A GPU Acceleration Library is a programming library that enables functions to be loaded onto the Graphics Processing Unit (GPU) memory to take the advantage of the superior parallel processing power of the GPU Memory Cores. Some GPU accelerated libraries used are:

- Nvidia CuDA
- Nvidia CuDNN
- Nvidia Rapids

GPU acceleration significantly reduces the code execution time hence improves the productivity and decreases the time to deployment.

2.1.6 Virtualization and Remote Access

Virtualization is the process of creating multiple virtual devices over a single system by use of an abstraction layer. Using virtual systems, one could access their workspace and in-house data from any authorized or supported device.

Remote access to the servers was performed via Secure Shell protocol (SSH) in order to access data or execute scripts. Data transfer from servers was performed using SCP (Secure Copy) protocol.

2.2 Technologies Summary

Table 1. Summary of Technological Stack

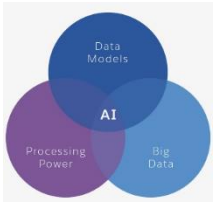
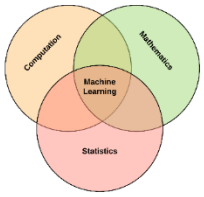
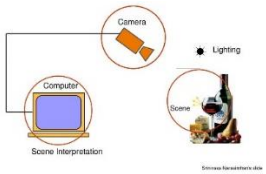
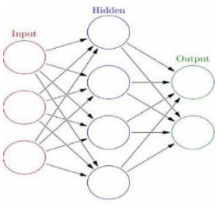
Sr No.	Image	Name	Use
1.	 **Fig 3.** AI Venn Diagram	Artificial Intelligence	• Backbone of products. • Create Programs that emulate human reasoning.
2.	 **Fig 4.** ML Venn Diagram	Machine Learning	• Create algorithms to process data. • Make predictions on data.
3.		Computer Vision	• Enable programs to percept their environment. • Key technology used by department.

	Fig 5. CV Descriptive Diagram		
4.	 **Fig 6.** DL Neural Network	Deep Learning	• Enable programs to learn from features in raw input. • Another key technology used alongside CV.
5.	NVIDIA CUDA **Fig 7.** Nvidia CUDA Library	GPU Acceleration Libraries	• Utilize GPU cores instead of processors to increase efficiency. • Use of high parallelism to reduce execution time.
6.	 **Fig 8.** Virtualization Diagram	Virtualization & Remote Access	• Provide virtual access to company systems and data. • Access servers using SSH and data using SCP protocols.

2.3 Hardware*

The Special Initiatives AI/ML department uses a range of hardware in order to generate code, create and deploy products, generate and gather data and train their Machine Learning and Deep Learning models. Some of the key hardware includes:

***Limited details provided to preserve company confidentiality**

2.3.1 Servers

The department is equipped with Ubuntu based server clusters decked with Intel Xeon Server Processors and Nvidia Tesla Architecture Graphics Processing Units. The servers are used to:

- Store in-house data.
- Test products.
- Train models.
- Data transfer.

The servers are remote accessed using authorized devices over a secure network using protocols such as SSH.

2.3.2 GPU Equipped Compute Systems

The department is also equipped with team allocated workstations which are decked with high performance Intel Xeon processors and Nvidia Quadro Graphics Processing Units. The base OS used on these workstations is Ubuntu Desktop LTS. Workstations are used for:

- Code generation.
- Model training and testing.
- Server access.

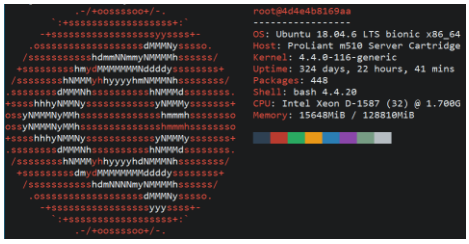
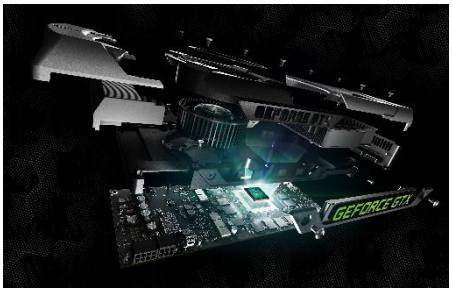
2.3.3 Computer Vision Sensors

The department uses their standardized and tested set of sensors for their Computer Vision related development and testing. Sensors are used as:

- Receptors for the product/system.
- Data input/generators.
- Product testing.

2.4 Hardware Summary

Table 2. Summary of Hardware

Sr No.	Image	Name	Use
1.	 **Fig 9.** Ubuntu Server	Servers	• Access in-house data and models. • Development and Testing. • Data transfer and storage.
2.	 **Fig 10.** Nvidia GPU Descriptive Image	GPU Equipped Compute Systems	• Code generation. • Model training. • Model testing. • Server Access.
3.	 **Fig 11.** Few CV Sensors	Computer Vision Sensors	• Used as receptors by program. • Used for program testing.

2.5 Programming Languages, Tools & Software

The New Initiatives AI/ML department at Jio Platforms uses a range of programming languages, frameworks, tools, and software to develop and implement its AI and ML-based products and services. Some of the key programming languages and tools used by the department include:

2.5.1 Python

Python is a high-level programming language that is widely used in the field of AI and ML. Python can be expanded via its various available open-source libraries.

The department develops its products and services using Python. Its simpler syntax, platform independence and availability of extensive libraries makes it one of the primary languages used by the department for training and testing ML models, analyzing and cleaning data, visualization of data etc.

Concepts of OOPs and modular code generation practices are followed in order to generate highly modular code, which is easy to integrate, test and debug. Best practices such as legible names, language case semantics, proper exception handling and assertion statements are used to develop code that can recover from any form of unexpected errors.

2.5.2 Python Libraries

The department uses the extensibility feature of python by installing python libraries for various tasks such as:

- Image Processing and Augmentation
    - OpenCV
    - Pillow
    - ImgAug
- Data analytics and visualizations
    - Matplotlib
    - Scikit Learn

- Seaborn

- Deep Learning Frameworks
  - PyTorch
  - TensorFlow
  - Fast.AI
- Data functions and Cleaning
  - NumPy
  - Pandas
- Directory functions
  - OS
  - Shutil

2.5.3 PyTorch

PyTorch is an open-source machine learning library that provides a seamless path from research to production. PyTorch is widely compatible and can be integrated within various ecosystems. PyTorch is used to define training and validation pipelines for model training and testing. Some use cases of PyTorch are:

- Defining model architecture.
- Loading and transforming data before training.
- Loading model.
- Training and exporting model.

2.5.4 Fast.AI

Fast.AI is a high-level abstraction layer using PyTorch as its backbone. Fast.AI makes training and optimizing AI ML models more accessible by simplifying the syntax, standardizing training pipelines and providing defined functions and utilities.

While the department is currently using PyTorch, Fast.AI is in the experimental phase of testing efficiency, reliability and scalability before being adopted and integrated within

the standard pipeline. The use case of Fast.AI are the same as PyTorch, some of its key features over PyTorch are:

- Simpler syntax.
- No boiler plate code required.
- Easy to migrate from PyTorch.
- Faster and more optimized.

However, it also has its own cons such as:

- Underexplained documentation.
- Lack of customizability due to abstraction.
- Smaller community support.

2.5.5 Jupyter

Jupyter Notebook is an open-source web application that allows the creation and sharing of documents that contain live code, equations, visualizations, and narrative text.

Jupyter Notebook kernels are used for data visualization and block by block code execution. The department uses Jupyter Notebook to write, run, and debug its code, as well as to share its findings with others.

2.5.6 IDE GUI

An Integrated Development Environment (IDE) is a software package consisting of tools such as:

- Code editor
- Code compiler
- Extensions and plugins
- Debugging utilities

The IDE used in the department is developer specific, however some of the most used IDEs are:

- Visual Studio Code
- JetBrains PyCharm
- Sublime text

2.5.7 IDE CUI

Console User Interface (CUI) IDEs are used when accessing servers using the SSH protocol to generate, test and debug code. Some of the most used CUI IDEs are:

- Vim
- Neo Vim
- Nano

IDEs such as vim and neo vim are highly customizable and work with a keyboard only philosophy, introducing a steep learning curve, but being much quicker, lightweight and efficient than the other IDEs.

2.5.8 Terminal Emulator

Terminal Emulators are applications which emulate traditional command line terminal. These terminals were primarily used in Ubuntu and when accessing a server. Terminals allow users to navigate through directories and execute commands using a command line interface. Some terminals used by the department were:

- Git bash
- Putty
- Tilix
- Gnome Terminal

2.5.9 Docker

Docker is an open platform for developing, shipping, and running applications. Docker enables one to separate their applications from their infrastructure so one can deliver software quickly.

It allows the containerization of Development Environments to run them independent of libraries and various dependencies. Docker is used by the department to containerize environments on remote servers for collaboration, testing and deployment. Some advantages provided by docker are:

- Platform and driver independence.
- Isolated dev environments on same machines.
- Avoid dependency conflicts.
- Have separate dev and prod containers.

Having docker containers allow developers to use the same machines simultaneously without having to worry about any dependency conflicts or environment pollution when installing new dependencies, packages or drivers

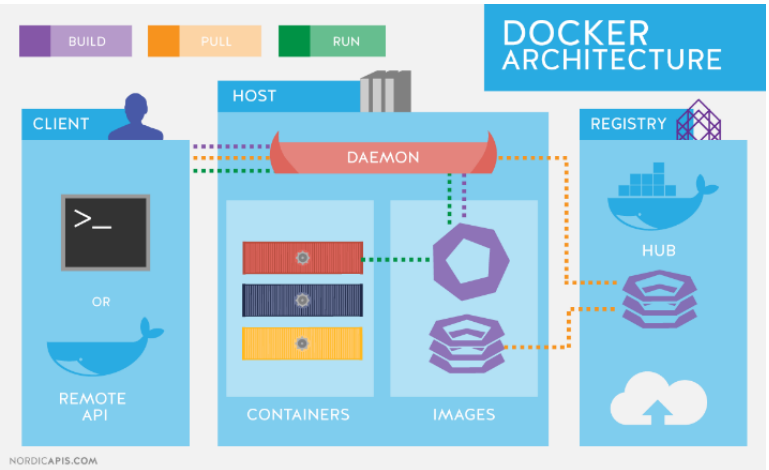


Fig 12. Docker Isolated Containers

The department is continuously exploring new and emerging technologies to stay at the forefront of AI and ML research and development.

2.6 Programming Languages, Tools & Software Summary

Table 3. Summary of Programming Languages, Tools & Software

Sr No.	Image	Name	Use
1.	 **Fig 13.** Python Logo	Python	• Simple syntax. • Extensible with libraries. • Main language used by department.
2.	 **Fig 14.** Commonly used libraries	Python Libraries	• Extending python. • Makes python a flexible multipurpose language.
3.	 **Fig 15.** PyTorch Logo	PyTorch	• Main framework for deep learning. • Used to create train/val pipelines. • Train and export models.
4.	 **Fig 16.** Fast.AI Logo	Fast.AI	• Simpler syntax compared to PyTorch. • More optimized and efficient.
5.	 **Fig 17.** Jupyter Logo	Jupyter	• Notebook format with code cell execution. • Used for visualization and collaboration.

6.	 **Fig 17.** VS code, a popular GUI IDE	IDE GUI	• GUI based code editor. • Used for code generation on workstations.
7.	 **Fig 18.** Neo Vim, a popular CUI IDE	IDE CUI	• CUI based code editor. • Used for code generation on servers.
8.	 **Fig 19.** Tilix, a popular Terminal	Terminal Emulator	• Used to access servers. • Command line interface
9.	 **Fig 20.** Docker Logo	Docker	• Containerize Dev environment. • Avoid dependency conflicts. • Platform independency.

Chapter 3 – Description of Projects

The New Initiatives AI/ML department at Jio Platforms is involved in several projects to develop and implement AI and ML-based products and services. Some of the key projects assigned by the department include:

3.1 Deep Learning Model for Facial Registration Quality Check

This project aims at developing a Deep Learning model to check the quality of the data captured by the facial registration and enrollment sensor to ensure that the extracted features are accurate, usable and were not subject to any form of artefacts which may lead to poor performance of the recognition system. Enrollment process is performed via a live camera feed sensor. A Deep Learning Development cycle was followed:

3.1.1 Problem Definition

The problem was defined as the need for a system that can ensure the quality of the facial registration data captured is accurate enough to ensure performance of the recognition system.

Usage of such poor-quality enrollment images can lead to an increase in the error rate and insult rate of the recognition system causing:

- Increased false positives.
- Increased false negatives.

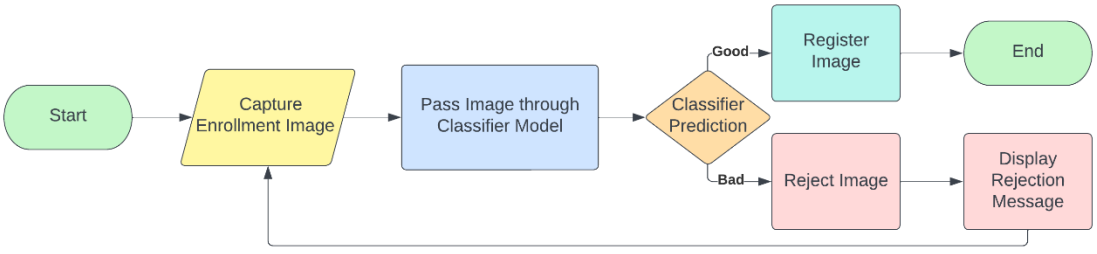


Fig 21. RQS Flow Diagram

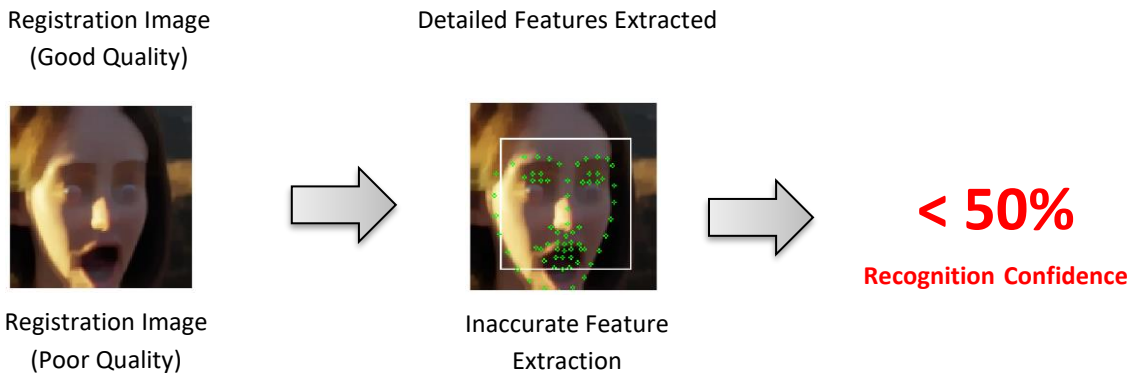


Fig 22. RQS Problem Diagram

3.1.2 Process Pipeline

1. **Data Collection:** Data collection involved the use of both open source and in house facial data to create a dataset having images labeled as good and low quality.
    a. Data from various sources was examined (open-source and in-house).
    b. Relevant data was compiled into a dataset till the data requirement was satisfied.
    c. The cumulated dataset folder was then randomly split into labels.
    d. Collected labelled data was split into a train/val split structure (80%/20%)

2. **Data Preparation:** The dataset created was pre-processed and image augmentation was performed to ensure robustness by introducing a more diverse range of data to the model.
    a. Image augmentation techniques were applied such as:
        i. Contrast manipulation
            - Sigmoid Contrast
        ii. Artificial blurring and Smoothening
            - Gaussian Blur
            - Motion Blur
            - Denoise Algorithm
        iii. Artificial Sharpening
        iv. Morphological Transforms etc

- Opening
- Closing

b. It was ensured each augmentation category was equally distributed in order to not have a biased dataset.

3. **Model Selection:** Fast.AI, a framework built over PyTorch was our primary deep learning framework. Pre-trained models were trained and tuned as per our criteria.
    a. Research regarding current best performing image classifier models was conducted.
    b. Using ImageNet dataset as a benchmark model architecture was selected to train and test.

4. **Model Training/Finetuning Pipeline:** Transfer learning pipeline was followed over pre-trained models where the dataset was loaded and the parameters were tuned to obtain a model checkpoint that performs as per our requirement.
    a. Relevant image processing models such as VGG, ResNet, ResNeXt were trained over custom compiled open-source dataset.
    b. Fitting data of the model training was plotted and analyzed.
    c. The hyperparameters (optimizer, learn rate, loss function) of the best checkpoints were tuned to further improve model performance.
    d. Best performing tuned models were selected for further fine tuning over in-house dataset.

5. **Model Evaluation:** An unseen test dataset was created to obtain the performance metrics of the model. Utilities such as GradCam were also used to further understand model activations.
    a. Model checkpoint was tested over unseen dataset.
    b. Metrics such as confusion matrix and F1 score were analyzed to conclude model performance.
    c. Utilities such as GradCam and Feature Ablation were implemented in order to determine mode layer activations and understand model performance and its generelization.

3.1.3 Process Pipeline Visualized

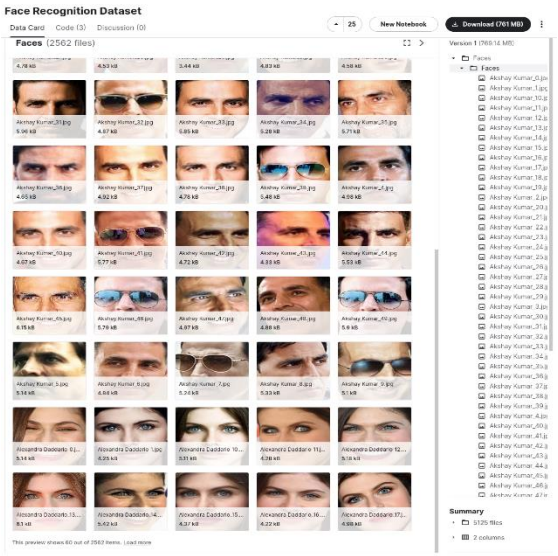


Fig 23. Dataset Compilation

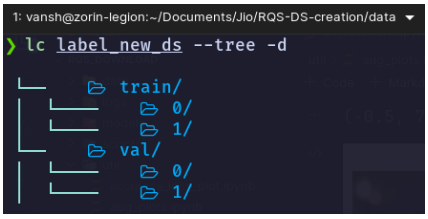


Fig 24. Train Val Split

Fig 25. Dataset Loaded as Dataloader

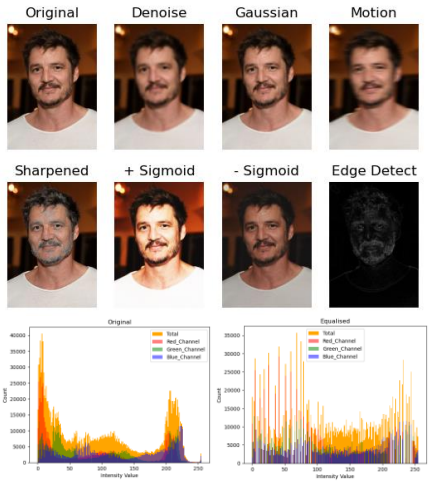


Fig 26. Image Augmentation

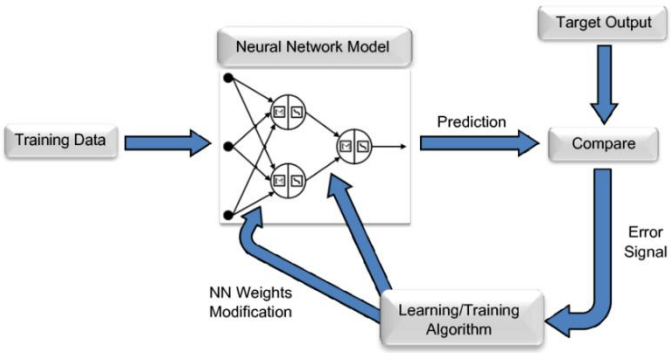


Fig 27. Model Training

		Condition Phase (Worst Case)		
		Condition Positive/ Shaded	Condition Negative/ Unshaded	
Testing Phase (Best Case)	Test Positive/ Shaded	True positive shaded T_p (*Correct*)	False positive shaded F_p (*Incorrect*)	Precision/Positive Predictive Value (PPV) $\frac{T_p}{T_p+F_p}$ x 100%
	Test Negative/ Unshaded	False negative unshaded F_n (*Incorrect*)	True negative unshaded T_n (*Correct*)	Negative Predictive Value (NPV) $\frac{T_n}{T_n+F_n}$ x 100
		Sensitivity/Recall Rate (RR) $\frac{T_p}{T_p+F_n}$ x 100%	Specificity Rate (SR) $\frac{T_n}{T_n+F_p}$ x 100%	

Fig 28. Model Evaluation (Confusion Matrix)

3.1.4 Technicalities Understood

1. **Transfer learning**
    a. Refers to the loading and tuning of pre-trained model architectures using pre-initialized weights.
    b. Transfer learning pipeline requires
        i. Initialization of the model
        ii. Loading dataset into a dataloader
        iii. Applying appropriate transforms
        iv. Loading the appropriate parameters
    c. Using transfer learning enables tuning of models to fit the use case required by the problem statement.

2. **Model fitting**
    a. Model fitting refers to the plots of the train and val acc and the train and val loss.
    b. Model fitting helps in identifying how well a model may perform on unseen data.
    c. There are 3 types of fitting
        i. Overfit – Model performs well on train and val data but performs poorly on unseen data. Caused due to poor generalization/bias.
        ii. Underfit – Model performs poorly on val data but performs well on train data. Caused due to lack of data or variety.
        iii. Good fit – Model can perform well on train val and test data.

3. **Good train accuracy does not guarantee good model**
    a. An overfitting model can indeed reach a train accuracy of 100%, however will not be able to perform on real world unseen data as it has not learnt enough generalization.
    b. Poorly fit models can still be finetuned and use as they may have learnt some form of feature extraction which can be used.

c. Making improvements to the dataset can also help improve model performance.

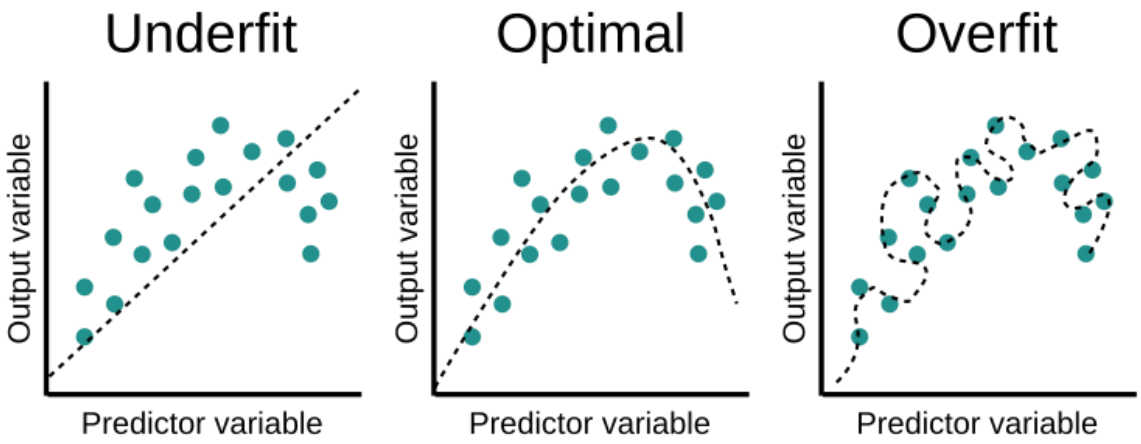


Fig 29. Model Fitting

4. **Good Dataset and Bad Dataset**
    a. It is nearly impossible to have a theoretically perfect database, however a good dataset refers to one which has enough variety of data for the model to identify and learn the underlying trend to have a good fitting.
    b. A bad dataset may have the following properties:
        i. Bias – Images don't have a lot of variety and may be similar to one another eg : Having majority images of a particular dog breed yields a model that performs poorly on other dog breeds.
        ii. Feature overload – Images may have too many features to identify the trend.
        iii. Not enough data – Quantity of data may fall short for model to identify generalization.
        iv. Poor train/val split – A dataset can be possibly split poorly leading to either set having a particular bias.

5. **Tuning Hyperparameters**
    a. Hyperparameters refer to model parameters such as loss function, optimizer, learn rate, number of epochs, decay, step size gamma etc.
    b. In general cases optimizers and loss functions are specified in the documentation.
    c. By analysing model train val data one can understand tuning parameters such as LR, decay, step etc.
    d. **Eg:** If a sudden drop in accuracy is seen after a few epochs of increase may indicate the minima was skipped due to high learning rate. In such cases the decay and LR need to be lowered in order to reach the minima.

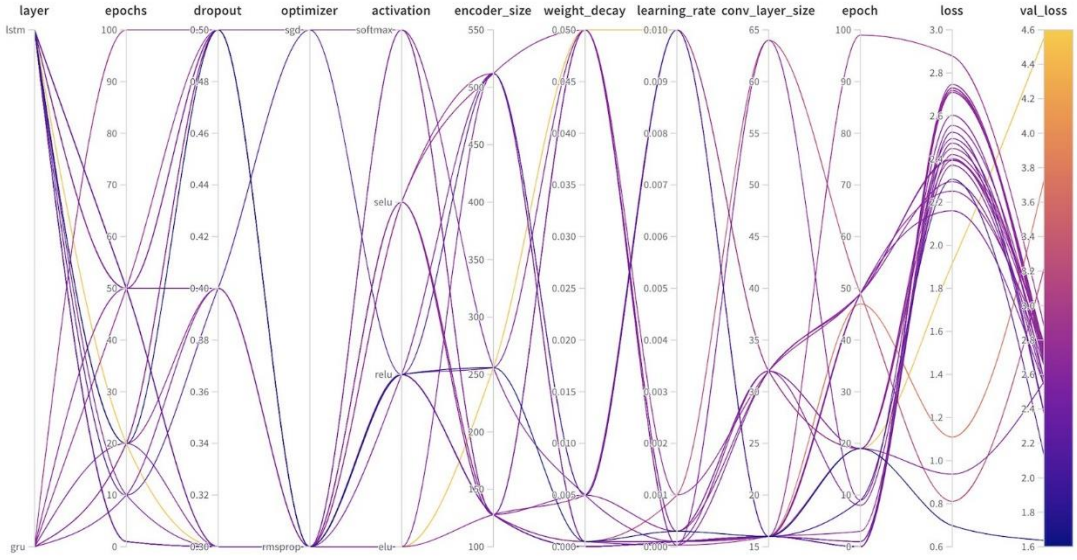


Fig 30. Hyperparameter Tuning

*3.1.5 RQS Progress and Results**

The given graph shows the F1score of each model.

Model nomenclature – **chkX-dsX-back**

1. **chkX –** Checkpoint number
2. **dsX –** Dataset name and number
3. **back –** Backbone model

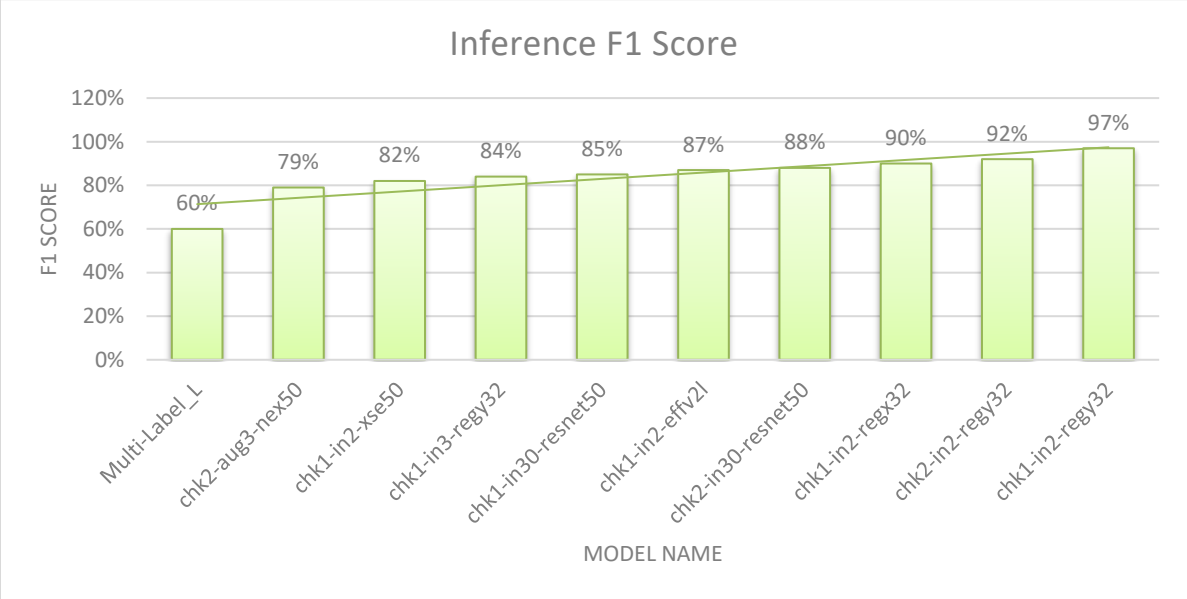


Fig 31. RQS Progress Plot

***Data such as confusion matrix not shown as per company confidentiality**

3.2 Spoof Detection Dataset Creation for Facial Recognition:

This project aims to develop a deep learning model for spoof detection in facial recognition systems. Spoof in this context is defined as the process of using a digital image, video or physical ID/printout of another person in order to bypass the implemented facial recognition system. The following steps were performed to create a training dataset:

3.2.1 Problem Definition

The problem was defined as the need for an accurate system that can recognize and flag an attempted spoofing of the system. Older checkpoint deployed has poor performance and yields many false positives, hence finetuning is required.

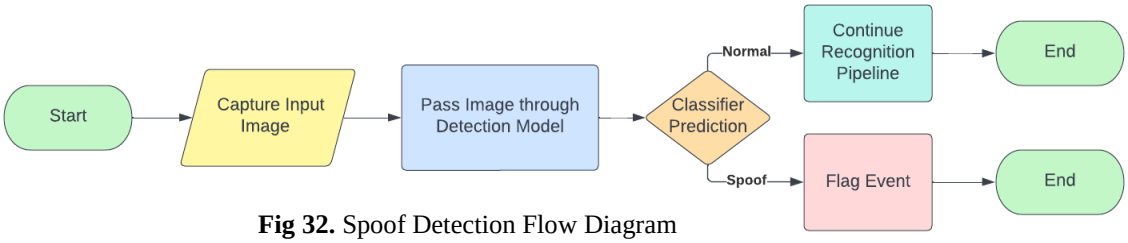


Fig 32. Spoof Detection Flow Diagram

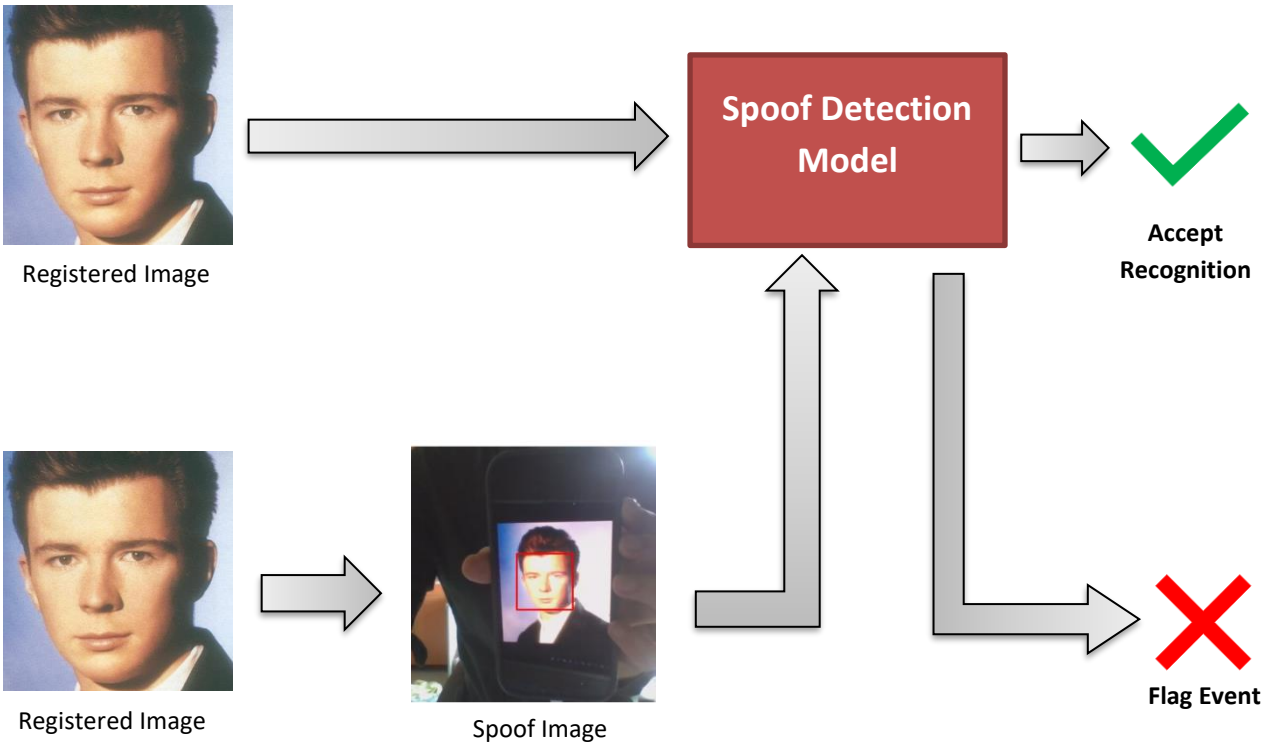


Fig 33. Spoof Detection Visualization

3.2.2 Process Pipeline

1. **Checkpoint Inferencing:** Images inferred by the older checkpoint were acquired and manual inferencing and utility inferencing were conducted.
    a. Checking if activations were strong on a particular feature.
    b. Checking if activations were random in similar images.
    c. Checking activations in images with same background to deduce a pattern.

2. **Data Collection:** Inhouse dataset was used in order to define images as spoof and False Positives to fine tune the detection model.
    a. Raw unlabelled data was received.
    b. Properties and features that may affect model performance were manually identified.
    c. Data was segregated with the assistance of image processing techniques and trained models into the identified labels.

3. **Data Preparation:** The dataset created consisted of cleaned inhouse data labelled as spoof and non spoof, the dataset was then pre-processed, augmented and used for finetuning model.
    a. Augmentation was applied to highlight certain features.
    b. Some features were removed from data by augmenting image.
    c. All images were cropped to a specific format before being used.

3.2.3 Process Pipeline Visualized

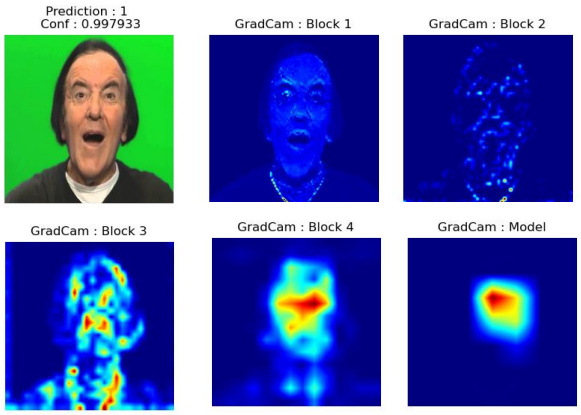
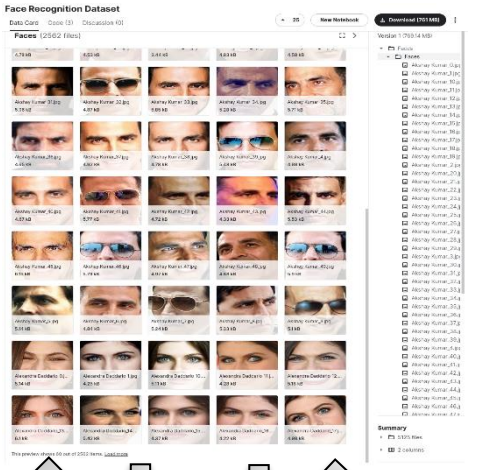


Fig 34. Checkpoint Inferencing with Gradcam



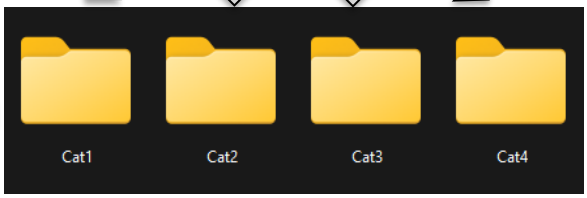


Fig 35. Segregate Raw Data in Categories

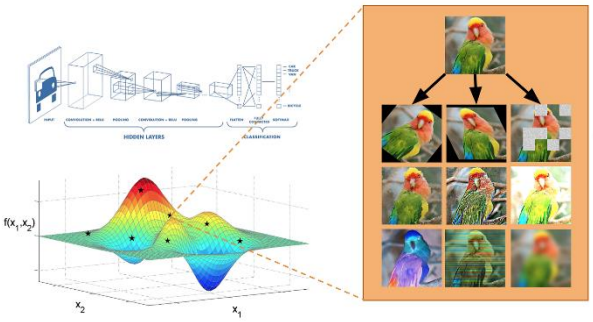


Fig 36. Perform Data Augmentation

Poor Performance

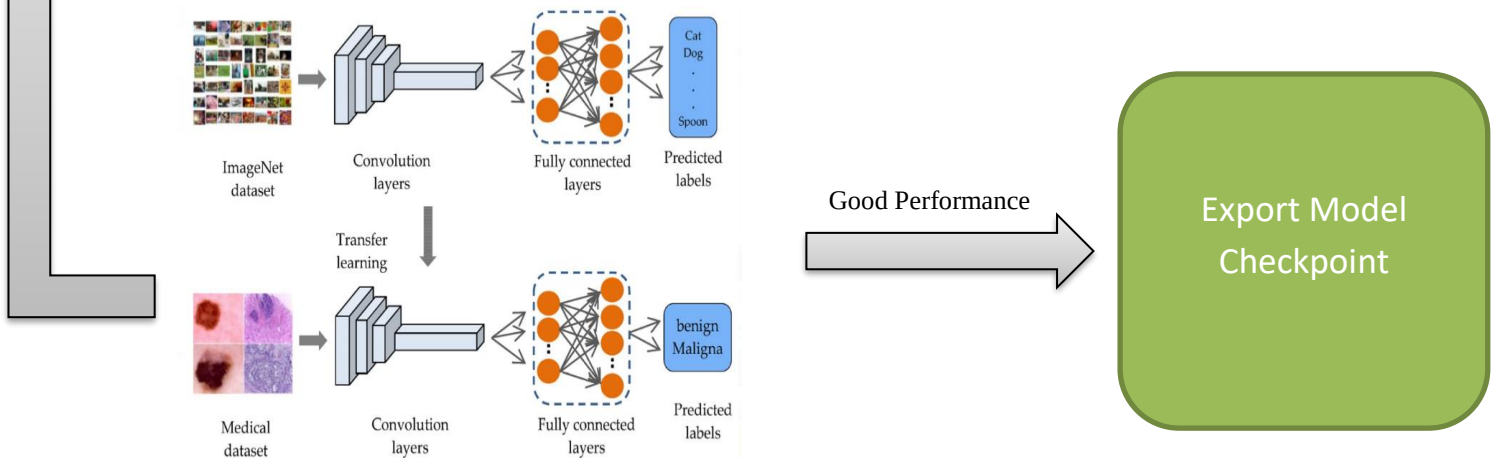


Fig 37. Fine tune Checkpoint

3.2.4 Technicalities Understood

1. **Data Cleaning**
    a. The process of removing noisy data and labelling data is called data cleaning.
    b. Data cleaning can be performed using assistive models and image processing scripts and techniques.
    c. To ensure a good dataset, manual intervention and effort is required as the scripts may not always label accurate data.
2. **Feature Overload**
    a. A model may experience feature overload when there are too many features in the image to extract and understand a trend.
    b. Data with feature overload may often lead to poorly fit models as the noise introduced by the data contributes to the model function negatively.
    c. Data with minimum to no feature overload yields the best models.
3. **Noisy Data**
    a. Data which has poor quality or not enough/too many features to extract is called noisy data.

b. Noisy data may affect the model function in a negative way, hence such data should be found and discarded.

Fig 38. Data with too many background features

Fig 39. Data with correct features

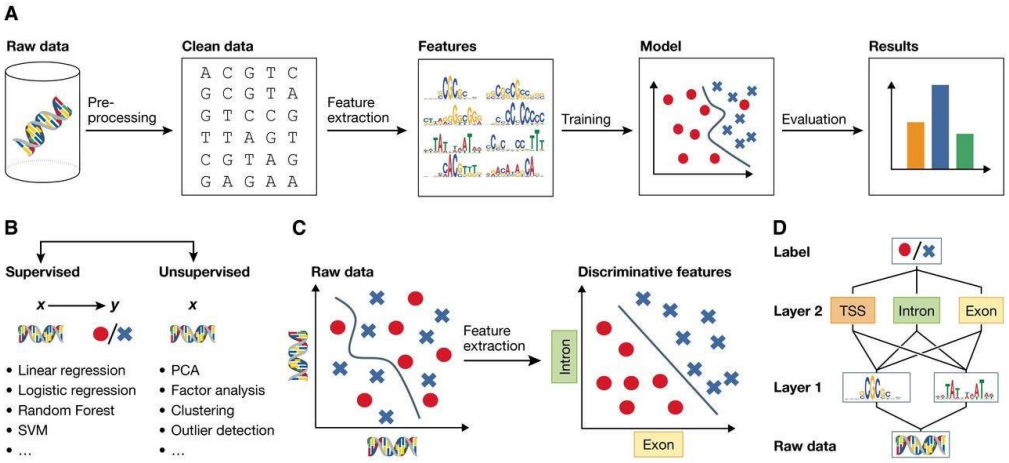


Fig 40. Model Training Tuning Pipeline

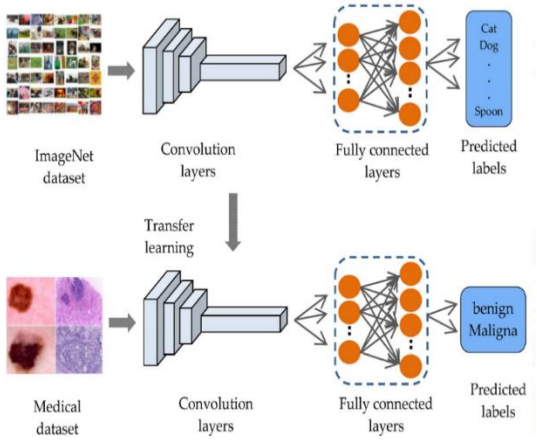


Fig 41. Transfer Learning

These projects demonstrate the department's commitment to developing and implementing innovative AI and ML-based products and services that drive business value for Jio Platforms.

*3.2.5 Spoof Progress and Results**

The given graph shows the F1score of each checkpoint and backbone.

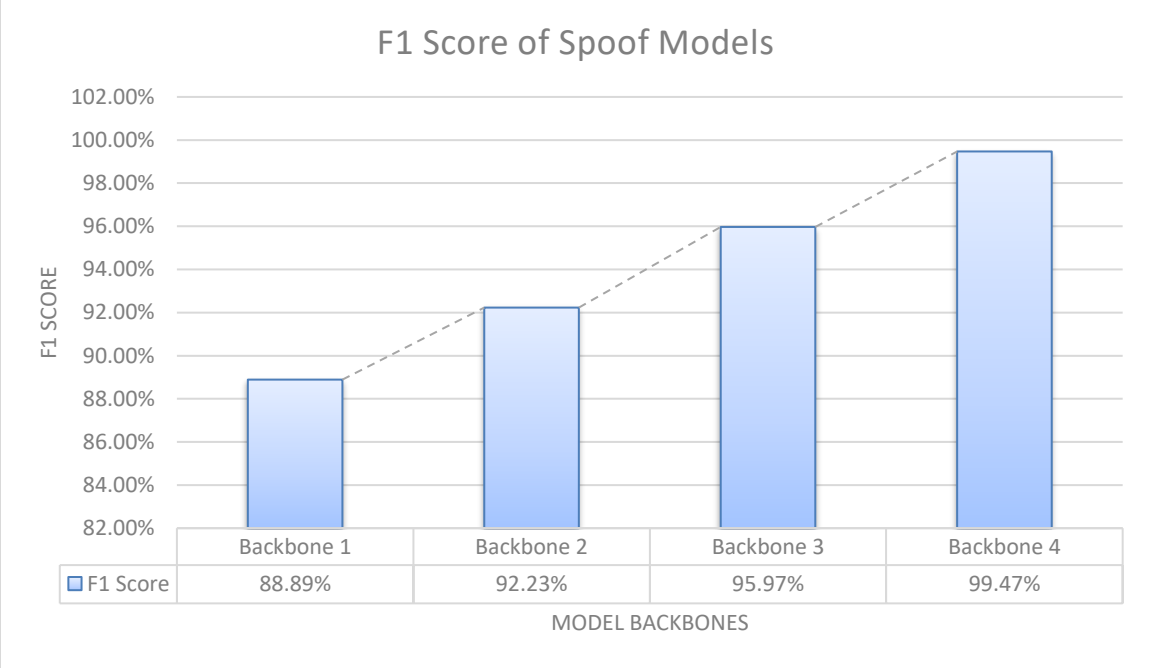


Fig 42. Spoof Detection Models Progress

***Data such as confusion matrix and backbone name not shown as per company confidentiality**

3.3 Research Projects:

3.3.1 Website Development Deck

This initial project involves conducting research and understanding the pipeline and flow of the website development process in order to identify what processes in the flow can be automated or assisted using AI.

Such automation and assistance can help improve developer productivity and efficiency as well as monitor and improve code and product quality. A presentation was created and was presented to our department superiors.

3.3.2 Software Factory Implementation

This project involves the development of a software factory that is designed to streamline the development and deployment of AI and ML-based products and services. The software factory is based on the Scaled Agile Framework (SAFe) and provides a standardized approach to software development, enabling the department to deliver high-quality products and services in a timely and efficient manner.

Thorough research was conducted in order to establish the architecture, flow and components of the software factory and a presentation of the topic was created and presented to our superior of the department.

3.3.3 Scaled Agile Framework (SAFe)

SAFe is a framework for scaling Agile development practices across multiple teams. The department has adopted the SAFe framework to ensure that its projects are delivered on time and with high quality, while also ensuring that teams are aligned and working together effectively.

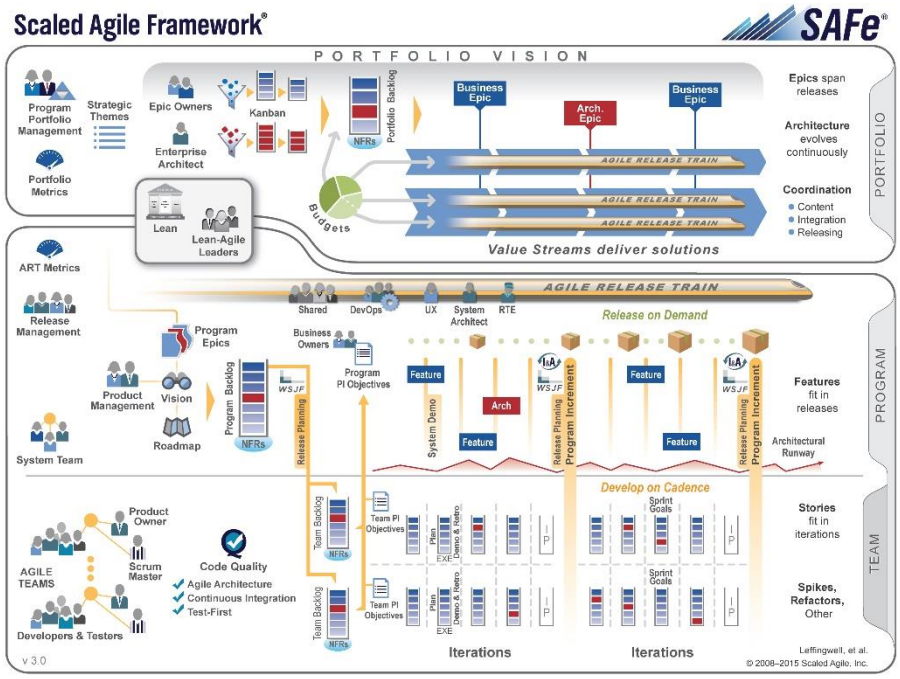


Fig 43. Scaled Agile Framework

3.4 Miscellaneous Projects:

3.4.1 Remote Sensor Onboarding

Sensor onboarding refers to the process of making installed sensors visible and accessible via inhouse software by use of specification configurations such as:

1. Admin Credentials
2. IP addresses

Once onboarded sensor data is accessible and recorded into the inhouse software.

3.4.2 Remote Sensor Certification

Sensor certification refers to the process of verifying the installation and functioning of the sensors and ensuring the sensors are performing as specified. Using our inhouse software interface sensor data was examined and sensors showing faults/anomalies/no recording were flagged. Criteria for unusual data was defined and faulty sensors were categorized as per the defined criteria.

Chapter 4 – Learnings and Deductions

4.1 Learnings in the Inplant Training

As an intern in the New Initiatives AI/ML department at Jio Platforms, I have had the opportunity to gain a wide range of technical skills and knowledge. Some of the key technical learnings that I experienced include:

4.1.1 Agile development methodologies

By working in an Agile development environment and following the Scaled Agile Framework (SAFe), I am learning how to work in a fast-paced, iterative development process. I also gained an understanding of how to work in a cross-functional team and how to deliver high-quality software on time.

4.1.2 AI and ML technologies

I have the opportunity to work with a range of AI and ML technologies, including deep learning and computer vision.

I am learning how to build and train models, how to evaluate their performance, and how to integrate them into real-world applications.

4.1.3 Software development best practices

I am learning about software development best practices, including code quality, testing, and continuous integration and delivery.

I am also gaining an understanding of how to work with source control systems such as Git, and how to use tools such as JioWorks and Kanban boards to manage software development projects.

4.1.4 Modular code generation

I am learning about modularizing my codes so that they can be reused in different applications and processes.

I have understood best coding practices and syntax and also understood how to write more developer friendly code using OOPs.

4.1.5 Data science and engineering

I am gaining an understanding of how to work with large datasets, how to clean and preprocess data, and how to perform exploratory data analysis. I am also learning how to use data visualization tools such as Matplotlib and Seaborn to visualize and understand your data.

4.1.6 Problem-solving skills

I have the opportunity to work on real-world projects, and as a result, I am developing problem-solving skills. I am learning how to identify and solve complex technical problems, and how to communicate my solutions effectively to stakeholders.
By working on these projects, I gained a solid foundation in AI and ML development and software engineering, which would be valuable for your future career. Additionally, I have the opportunity to network with experienced professionals in the field, which could lead to future opportunities for growth and advancement.

Chapter 5 – Conclusion

Conclusion and Future Plan of Action

As an intern in the New Initiatives AI/ML department at Jio Platforms, I have the opportunity to work on exciting projects and gain valuable experience in the field of AI and ML. In conclusion, I would likely have the following takeaways and future plans of action:

Valuable technical skills and knowledge

I have gained a solid foundation in AI and ML development, software engineering, and data science, which would be valuable for your future career.

Networking opportunities

I have had the opportunity to network with experienced professionals in the field, which could lead to future opportunities for growth and advancement.

Passion for AI and ML

My experience in the New Initiatives AI/ML department has sparked and further solidified my passion for AI and ML, and I have a desire to continue learning and exploring these exciting fields.

Based on these experiences, my future plan of action includes:

Building my portfolio

I would want to continue building my portfolio of AI and ML projects, both through my work at Jio Platforms and through personal projects.

Networking with professionals in the field

I would want to continue building my network of professionals in the field, through online communities, professional organizations, and conferences.

Staying up to date with the latest developments

I would want to stay up to date with the latest developments in AI and ML, through reading industry publications, attending conferences, and participating in online communities.

In conclusion, my experience as an intern in the New Initiatives AI/ML department at Jio Platforms is a valuable steppingstone for my future career, and my future plans of action would likely reflect your passion for AI and ML and my desire to continue learning and growing in this exciting field.

CERTIFICATE OF INPLANT TRAINING

This is to certify that Mr./~~Ms~~ Vansh Shah SAP ID. 70321019094 of **(VIII/~~ XI~~) Semester** of B.Tech Integrated Program Computer ~~/EXTC/Civil/Mechanical~~ has attended the Inplant Training for 83 days out of 83 at Jio Platforms Limited (JPL) during the Period: from 2nd January 2023 to 28th April 2023.

Signature of the Student

Signature of the Industry Mentor

Name: Aditya Chauhan
Email ID: aditya3.chauhan@ril.com
Contact: +91 9326222811
Date: 04/05/2023

Forwarded with compliments to the HoD for necessary action.

The details of working at organization:

Present: 83 Days
Leave: 0 Days
Absent: 0 Days
Total: 83 Days