**DEVELOPING A MACHINE LEARNING-BASED TOOL FOR IDENTIFYING PLANT DISEASES**

**Dr. Chandra Shekhar Tyagi**

[chandrah@srmist.edu.in](mailto:chandrah@srmist.edu.in)

ADITYA BHARDWAJ [ab2225@srmist.edu.in](mailto:ab2225@srmist.edu.in)

YASH SINGHWAL [ys1394@srmist.edu.in](mailto:ys1394@srmist.edu.in)

DEEPANSHU [du4970@srmist.edu.in](mailto:du4970@srmist.edu.in)

PARAS RAJPAL [pr1524@srmist.edu.in](mailto:pr1524@srmist.edu.in)

Department of CSE

SRM NCR Campus Modi Nagar

April 2023

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ABSTRA****CT**

Our project's objective is to develop a machine learning model that effectively classifies plant diseases using image classification techniques. Our primary objective is to address the crucial need for timely and precise identification of plant diseases, as they have a significant impact on crop yields and food security.

Our project’s objectives include training a model capable of accurately recognizing different types of plant diseases from images and creating a user-friendly interface for farmers and agricultural professionals to use in the field. We will utilize a dataset comprising images of both healthy and diseased plants. We will employ convolutional neural networks, a form of deep learning, to classify the images and detect plant diseases. The model will be trained and evaluated using transfer learning, data augmentation, and k-fold cross-validation techniques.

We expect our machine learning model to exhibit a high level of accuracy in the detection of plant diseases and demonstrate robust generalization capabilities when applied to new images. Additionally, we expect that the user-friendly interface will make the model practical and accessible for farmers and agricultural professionals.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**INTRODUCTION**

The detection of plant diseases is critical in ensuring the health and productivity of crops. Automatic techniques for detecting plant diseases have become increasingly important in monitoring large farms and alerting producers to the appearance of diseases on plant leaves. Ensuring timely and precise detection of plant diseases is of utmost importance, particularly considering the growing global population and the necessity to cultivate sufficient crops to sustain the world's food needs.

Unfortunately, plants can be affected by various diseases that cause a significant loss in productivity. These diseases can result in decreased yield, poor quality crops, and an increased risk of contamination. Plant diseases can also spread quickly and affect large areas of crops, leading to devastating consequences for farmers and the food supply.

To address these challenges, we can leverage a combination of technologies such as artificial intelligence, machine learning and convolutional neural networks (CNN). By employing these cutting-edge technologies, we can enhance our capabilities in combating these issues effectively. By implementing advanced techniques, we can provide smarter solutions for detecting and identifying plant diseases.

Plant disease detection falls under the category of predictive modelling, where the goal is to predict a class label for a given input data sample. In this case, the output of the model will be binary, indicating either "yes" or "no" to signify whether the plant is diseased or healthy, respectively.

The application of artificial intelligence and machine learning techniques in plant disease detection has proven to be highly effective. By training machine learning models on large datasets containing images of healthy and diseased plants, these models can learn to recognize and distinguish the visual patterns associated with different types of plant diseases.

Convolutional neural networks (CNNs) have become widely used and highly effective in the field of plant disease detection. This deep learning technique excels at extracting significant image features and using them for accurate classification. The training process involves exposing the networks to extensive collections of images, enabling them to learn and generalize from the data. As a result, the trained models demonstrate impressive accuracy when classifying new images.

Our project aims to create an accurate machine learning model for identifying and classifying plant diseases using image classification techniques. We also aim to develop a user-friendly interface for easy field identification of diseased plants.

In conclusion, the detection of plant diseases is essential for ensuring the health and productivity of crops, and automatic techniques can greatly assist in this process. By utilizing advanced technologies, we can develop smarter solutions for detecting and identifying plant diseases. With the increasing global population and the need to grow enough crops to feed the world, it is more important than ever to develop effective and efficient methods for monitoring and maintaining the health of crops.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**LITERATURE SURVEY**

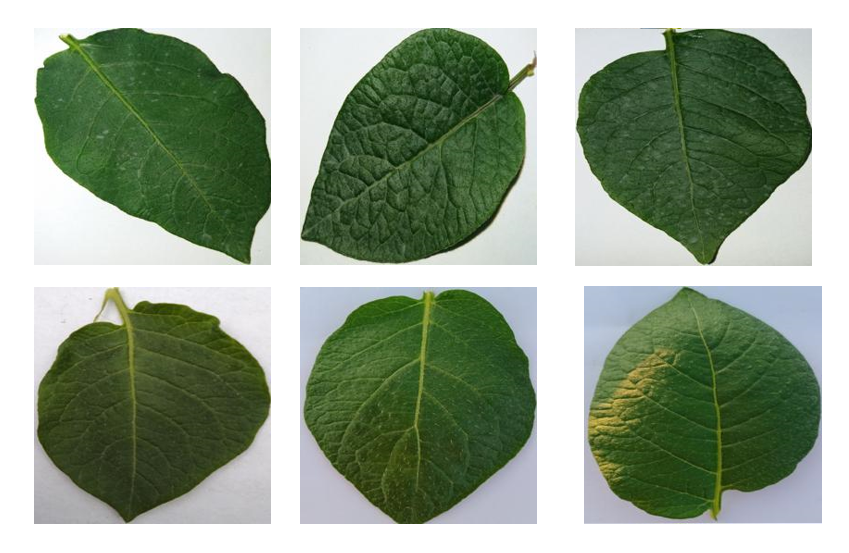
|  |  |  |
| --- | --- | --- |
| **Paper and Author** | **Methodology** | **Conclusion** |
| “Detection of Leaf Disease and Classification Using Digital Image Processing” (2012) by RM Prakash | Support vector machines (SVMs) and k-means clustering | The proposed approach achieved an impressive classification accuracy of 92.2% when distinguishing various plant diseases. |
| “Applying Image Processing Techniques to Detect Plant Diseases” (2012) by Kulkarni et al | Artificial neural networks (ANNs) | The proposed technique exhibited a detection accuracy of 91% when identifying plant diseases. |
| “Detection of Unhealthy Region of Plant Leaves and Classification of Plant Leaf Diseases Using Texture Features” (2013) by Arivazhagan Selvaraj | Color transformation, segmentation, and classification | With an accuracy of 90%, the proposed method successfully identified unhealthy regions on plant leaves. |
| “Leaf Disease Detection and Classification Using Artificial Neural Network” (2015) by Malavika Ranjan, Manasi Rajiv Weginwar, and A. Ingole. | Artificial neural networks (ANNs) | ANNs were trained to distinguish diseased plants from healthy samples by learning a set of features that were carefully chosen. Model successfully achieved an accuracy of 85%. |

|  |  |  |
| --- | --- | --- |
| **Paper and Author** | **Methodology** | **Conclusion** |
| “Deep Learning for Image-based Plant Detection” (2016) by Sharada P. Mohanty and David P. Hughes. | Convolutional Neural Network (CNNs) | The model underwent training to differentiate between healthy and diseased plants across fourteen distinct species, achieving an exceptional accuracy of 99% on the test set. Although this performance exceeds random selection, incorporating a broader range of training data can lead to further improvements in accuracy. |
| “Cucumber Disease Detection Using Artificial Neural Network” (2016) by P. Pawar | Artificial neural networks (ANNs) | The utilization of supplementary texture features can enhance the classification accuracy. |
| “Detection and Measurement of Paddy Leaf Disease Symptoms Using Image Processing” (2017) by RP Narmadha | Artificial neural networks (ANNs) | The study assesses multiple image processing techniques used in the detection and diagnosis of crop leaf diseases. |
| “Convolutional Neural Network-based Inception v3 Model for Animal Classification” (2020) by Jyotsna Bankar | Convolutional Neural Network (CNNs) | The model is not only useful for object classification but also for categorization, rendering it an asset in diverse image classification tasks. |
| “Plant Disease Detection Using CNN and GAN” (2022) by Emanuel Cortes | Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) | Accurate feature extraction and output mapping heavily rely on background segmentation, which plays a critical role in various applications. Intriguingly, the utilization of Generative Adversarial Networks (GANs) demonstrates promising prospects in disease classification for plants. |

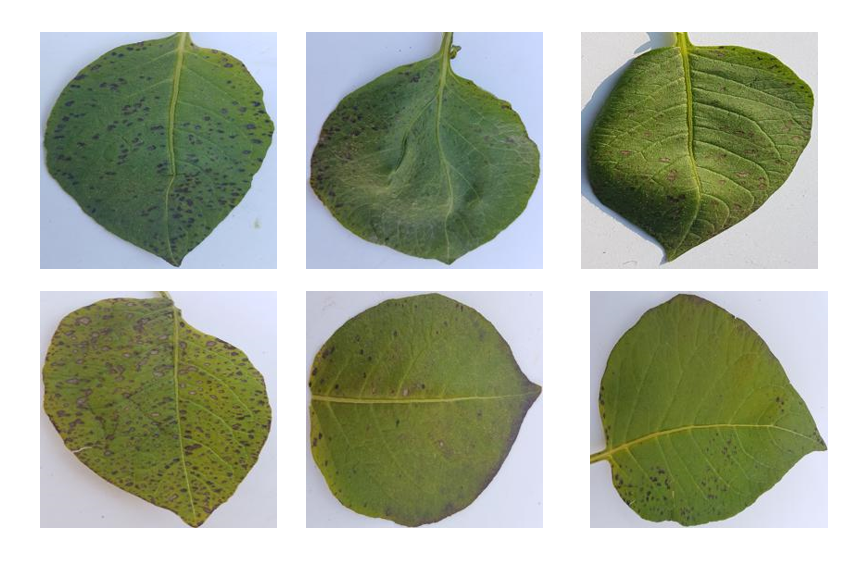
**DATASET/ DATABASE SPECIFICATION**

The main characteristics of the training and testing datasets employed in this study encompass the following:

* + Both the training and testing datasets consist of 2152 high-quality, real-life symptom images depicting various potato diseases.
  + The images in the dataset are classified into three distinct groups: Early blight, Late blight, and healthy.
  + Potato Early blight is attributed to the fungus Alternaria solani and manifests symptoms like leaf spots and discoloration.
  + Potato Late blight is induced by the oomycete pathogen Phytophthora infectants and can infect potato foliage and tubers at any stage of crop development. This disease can have devastating effects on crop yields.
  + The Potato healthy group represents images of healthy potato leaves that are free from any disease.
  + The training set was used to train the machine learning model, while the testing set was utilized to evaluate the model's accuracy.
  + The inclusion of high-quality images in the dataset is crucial for ensuring precise and dependable disease diagnosis.
  + By training the machine learning model on a diverse set of images, we can im- prove its ability to accurately identify different types of potato disease.



**Figure:** Images depicting Healthy Potato



**Figure:** Images depicting Early blight Potato



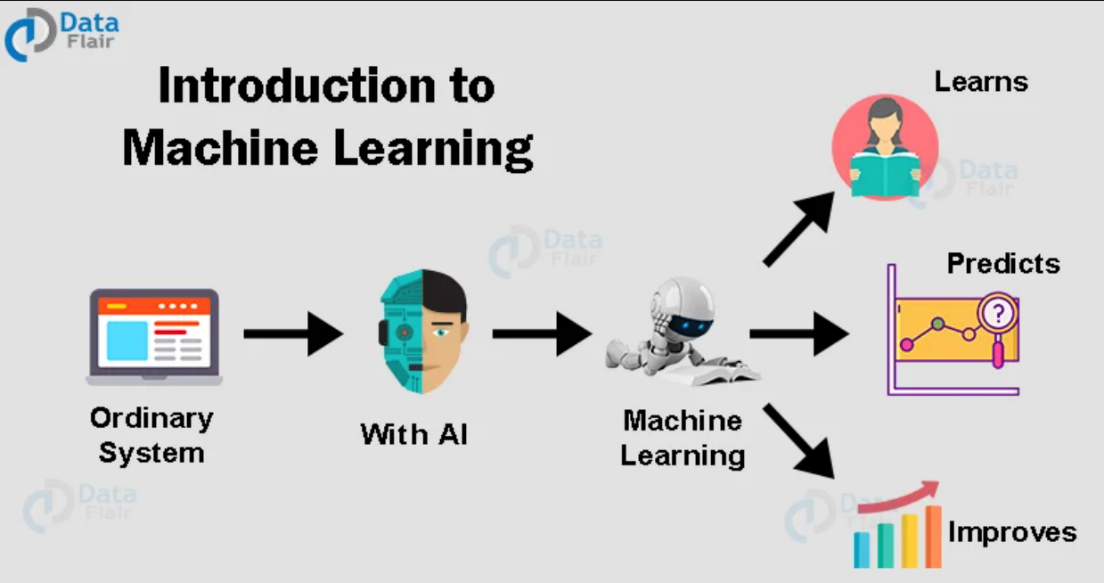
**Figure:** Images depicting Late blight Potato

**METHODOLOGY**

The technology used are: -

1. **Machine Learning:**
   * Machine learning, a field of artificial intelligence, utilizes algorithms to analyse and extract knowledge from data.
   * It empowers software applications to make accurate predictions without explicit programming
   * Machine learning algorithms leverage past data as input to generate new output values.
   * It enables businesses to uncover patterns in customer behaviour and operational trends, facilitating the creation of innovative products.
   * Prominent companies such as Facebook, Google, and Uber integrate machine learning extensively into their operations to gain a competitive edge

In summary, machine learning is a vital tool for businesses looking to stay ahead of the competition. By analysing data and generating accurate predictions, it allows companies to develop new products, track customer behaviour, and gain valuable insights into operational patterns. The technology’s potential for making sense of big data has made it an essential tool for businesses across various sectors, and it is expected to continue growing in importance in the years ahead.



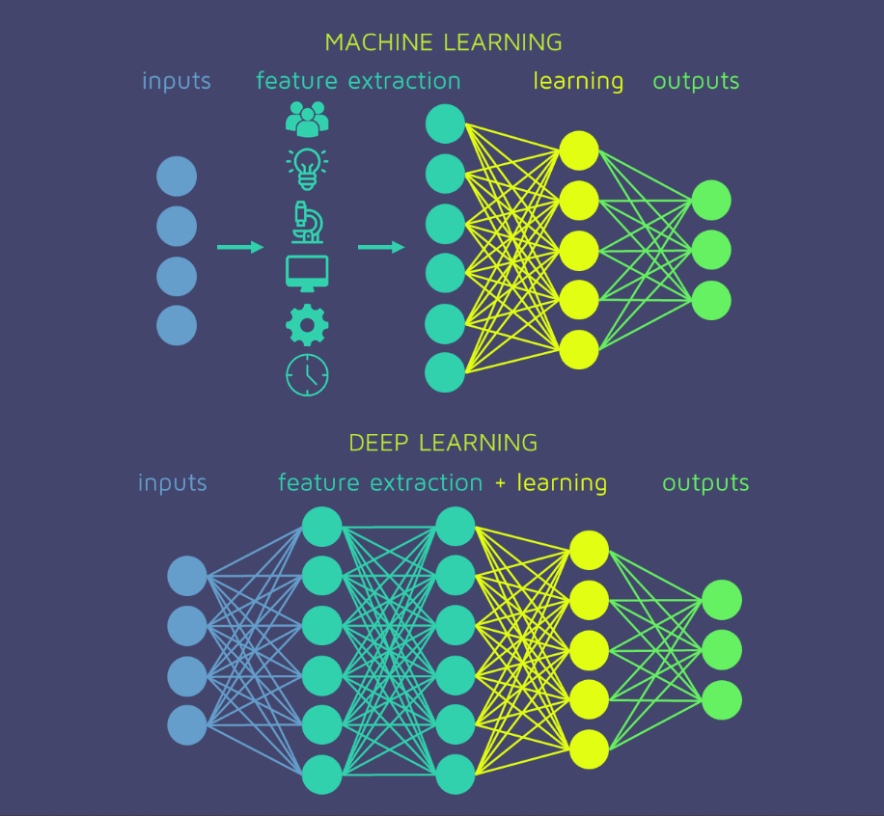
**Figure:** Machine Learning

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**2. Deep Learning:**

* Deep learning is a branch of machine learning that leverages neural networks to extract insights from data and generate accurate predictions.
* It draws inspiration from the structure and functionality of the human brain, featuring interconnected nodes organized into multiple layers for information processing.
* Deep learning models undergo training on extensive datasets, and their accuracy improves with the volume of training data.
* It finds applications in various fields, including speech recognition, computer vision, and natural language processing.
* Deep learning has led to remarkable advancements in areas such as autonomous vehicles, speech recognition, language translation and image recognition.
* Effective training of deep learning models necessitates substantial computational power and data, which can pose challenges for smaller organizations.

In summary, deep learning harnesses the power of neural networks to learn from data and make predictions. It has brought about revolutionary changes in fields like computer vision and natural language processing, holding the potential to transform numerous industries. Nonetheless, the effective training of deep learning models demands substantial computational resources and data, which can pose challenges for certain organizations.



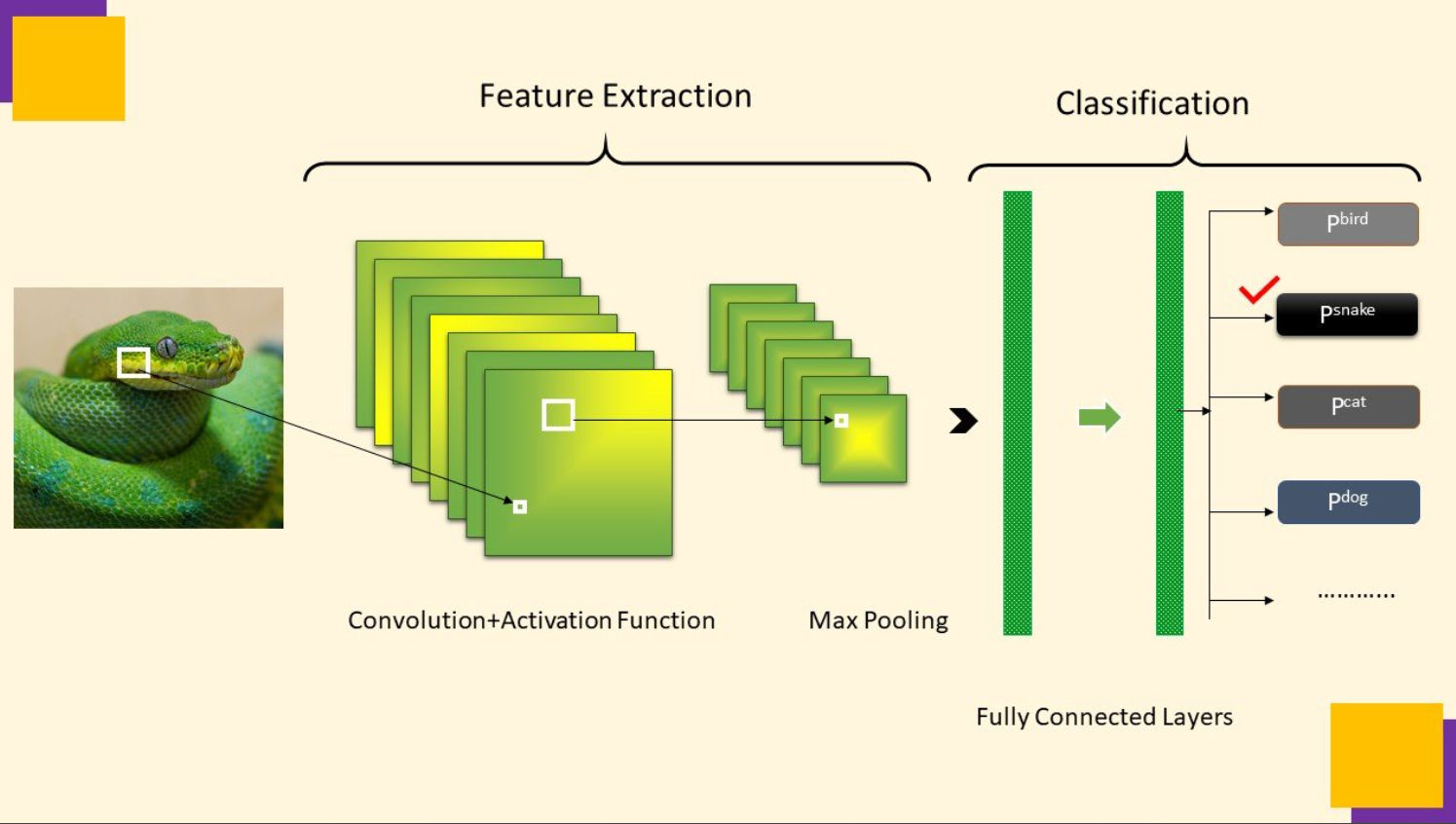
**Figure:** Deep Learning

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**3. Convolutional Neural Network (CNN):**

* Convolutional Neural Networks (CNNs) are a type of deep neural networks that are specifically designed for tasks that involve image and video recognition.
* CNNs draw inspiration from the structural organization of the human brain's visual cortex, utilizing filters that extract image features at various scales and orientations.
* CNNs employ a technique called convolution, where a filter slides across an image and multiplies its values with the corresponding pixel values in the image.
* CNNs can learn to recognize complex patterns in images and classify them into different categories.
* They have been used in applications such as self-driving cars, facial recognition, and medical imaging.
* Training a CNN requires a large dataset of labelled images, which can be a challenge for some applications.

In summary, Convolutional Neural Networks (CNNs) are deep neural networks specifically developed for image and video recognition tasks. They draw inspiration from the human visual cortex's structure and utilize convolution to extract features from images. CNNs are powerful tools for recognizing complex patterns in images and have applications in fields like self-driving cars and medical imaging. However, they require a large dataset of labelled images to be trained effectively, which can be a challenge for some applications.



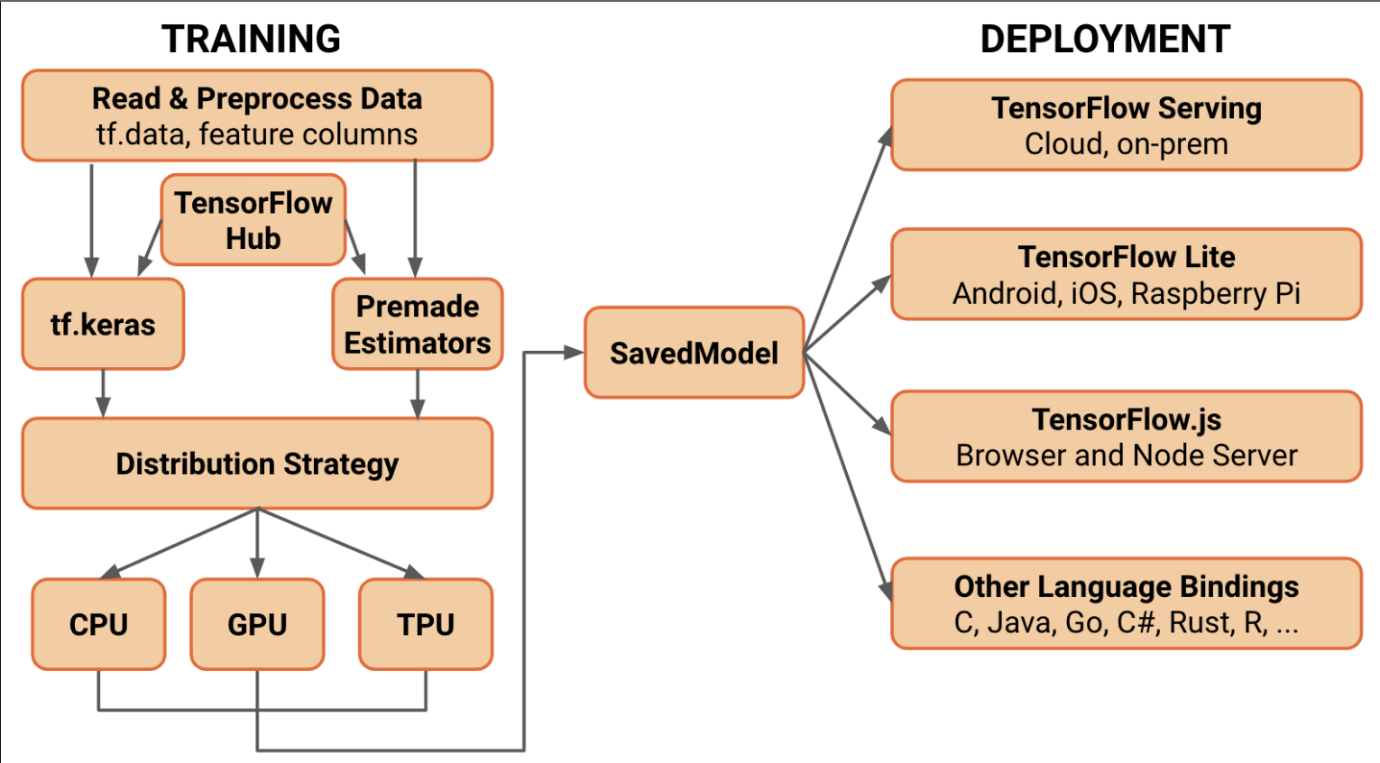
**Figure:** Convolutional Neural Network

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**4. Sequential Model:**

* Sequential models represent a class of neural networks primarily employed for processing sequential data, including time series and natural language.
* They consist of interconnected layers of nodes that sequentially process input data.
* A common example of a sequential model is the Recurrent Neural Network, which incorporates a "memory" to process inputs in relation to previous inputs.
* Another variant is the Long Short-Term Memory network, designed to address the issue of vanishing gradients that may arise in RNNs.
* Sequential models find utility in domains like speech recognition, text generation, and sentiment analysis.
* Training a sequential model necessitates a substantial dataset of labelled sequential data, which can pose challenges for certain applications.

In summary, Sequential Models are neural networks specifically designed for processing sequential data. They find applications in domains like speech recognition and natural language processing. RNNs and LSTMs are two notable types of sequential models that excel at handling sequential data. However, training sequential models requires a large dataset of labelled sequential data, which can be a challenge for some applications.



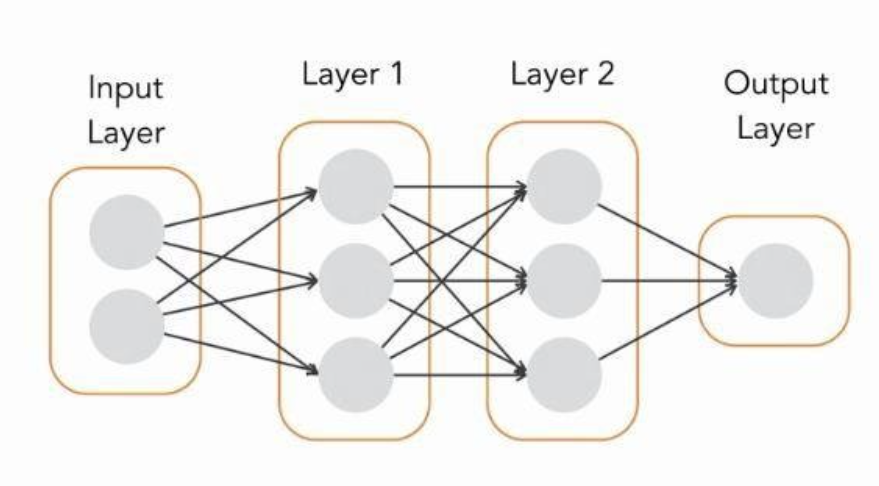
**Figure:** Sequential Model

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**5. TensorFlow:**

* TensorFlow is a freely available software library used for constructing and training machine learning models. It is an open-source framework.
* Created by Google, TensorFlow is widely recognized as one of the most extensively utilized machine learning library.
* It provides a wide range of tools for building neural networks and other ma- chine learning models, including tools for data pre-processing, model building, and model evaluation.
* TensorFlow supports a wide range of platforms, including CPUs, GPUs, and mobile devices.
* It allows users to build models using a high-level interface (such as Keras) or a low-level interface for maximum flexibility and control.
* TensorFlow finds applications in various domains, including predictive analytics, natural language processing, and image and speech recognition.

In summary, TensorFlow is a renowned machine learning library extensively employed for constructing machine learning models. It offers a comprehensive set of tools for model development and evaluation, supporting various platforms. It enjoys widespread adoption in both research and industry, playing a pivotal role in driving numerous recent advancements in machine learning.



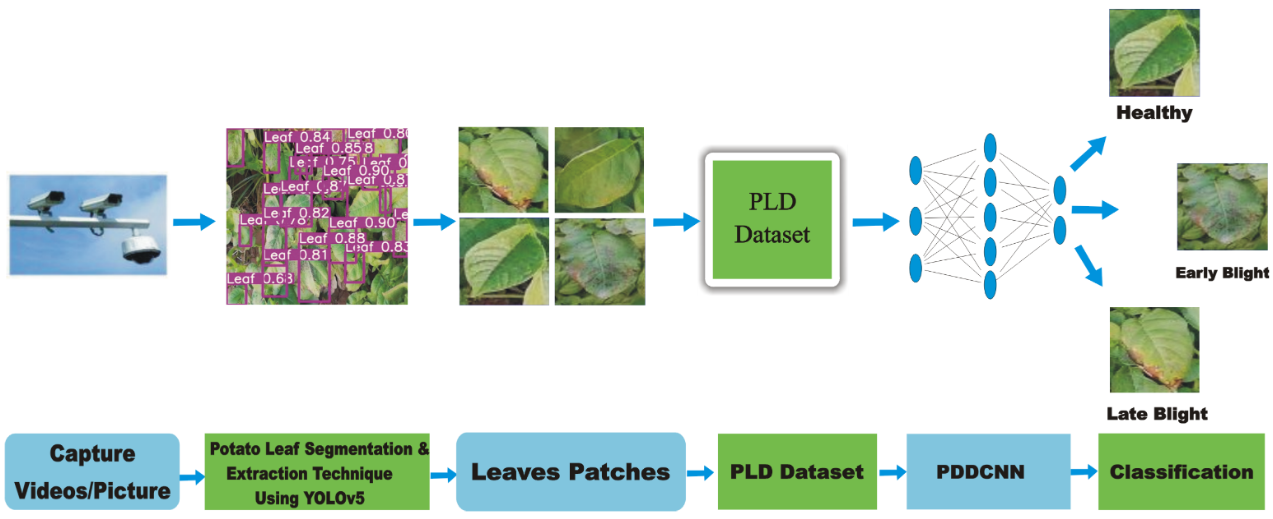
**Figure:** TensorFlow

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**6. Image processing:** Image Processing is the use of algorithms to transform and analyse images. Image processing encompasses a diverse range of techniques, each with distinct applications and benefits.

Here are five main types of image processing:

* **Visualization**: Visualization techniques are used to enhance the visual appearance of images to improve human perception. This includes techniques like contrast stretching, histogram equalization, and colour mapping.
* **Recognition**: Recognition techniques are used to identify objects or features in an image. These techniques encompass various aspects such as facial detection, optical character recognition (OCR) and object detection.
* **Sharpening** **and** **Restoration**: Sharpening and restoration techniques are used to improve the quality of images that are degraded by noise or other factors. This includes techniques like noise reduction, deblurring, and super- resolution.
* **Pattern** **Recognition**: Pattern recognition techniques are used to identify pat- terns or textures in an image. This includes techniques like texture analysis, edge detection, and feature extraction.
* **Retrieval**: Retrieval techniques play a crucial role in searching for images that exhibit similarity to a given image or set of images. These techniques include content-based image retrieval (CBIR) and image classification.

In summary, Image Processing is a broad field with many different types of techniques used to transform and analyse images. The five main types include visualization, recognition, sharpening and restoration, pattern recognition, and retrieval. Each type of image processing technique possesses its own distinct applications and methods, finding utility in diverse fields such as medicine, manufacturing, and security.

**Figure:** Image processing

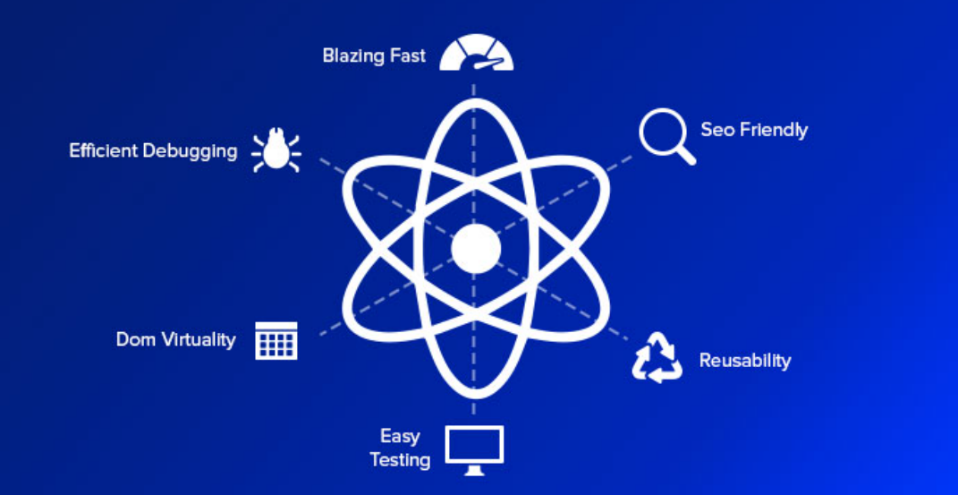
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**7. ReactJS:** ReactJS is a well-known JavaScript library that is widely used for developing user interfaces in an open-source manner. Originally developed by Facebook, ReactJS is now embraced by numerous companies, regardless of their size. It empowers developers to create intricate and interactive user interfaces effortlessly. ReactJS simplifies the process of state management within web applications, facilitating efficient development practices.

Some key features of ReactJS includes:

* **Component-Based Architecture:** ReactJS revolves around the concept of decomposing an application into reusable and modular components. This approach simplifies the development and maintenance of intricate user interfaces, allowing for efficient construction and management.
* **Virtual DOM:** To enhance performance and minimize the need for excessive updates during rendering, ReactJS employs a virtual representation of the Document Object Model (DOM). By utilizing this virtual DOM, ReactJS optimizes the efficiency of rendering changes in the user interface.
* **One-way data binding**: ReactJS follows a unidirectional data flow, which implies that modifications made in the user interface do not directly impact the application state. This unidirectional flow simplifies the management of complex state changes, enhancing the overall control and predictability of the application's behaviour.
* **JSX**: ReactJS utilizes JSX, a syntax that enables developers to write HTML-like code within JavaScript. This integration of HTML-like syntax with JavaScript facilitates a more comprehensible and manageable structure for web applications, streamlining the development process.

In summary, ReactJS is a popular JavaScript library for building interactive user interfaces. Its component-based architecture and efficient handling of the virtual DOM make it a powerful tool for creating complex web applications. It has a large and active community and is often used in combination with other technologies to build web and mobile applications.



**Figure:** ReactJS

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**CONCLUSION**

In conclusion, the utilization of machine learning techniques, including Convolutional Neural Networks (CNN), Keras, Sequential Models and Tensorflow, has demonstrated effectiveness in plant disease diagnosis. This study specifically focused on processing and segmenting images of potato leaves infected with Early Blight, Late Blight, and healthy potato. The results showed a significant improvement in accuracy, from 90% to 95%, which indicates that the method is reliable for diagnosing plant diseases.

However, it is crucial to recognize that this approach requires investigation to address these challenges with this approach. One such limitation is the requirement for extensive labelled data to effectively train the model. Another limitation is the difficulty of distinguishing between different types of diseases that may have similar symptoms.

Despite these limitations, the potential of machine learning for plant disease diagnosis is significant. By combining advanced imaging technology with machine learning algorithms, we can accurately diagnose diseases early, leading to improved crop yield and reduced economic losses. In summary, this study provides a notable contribution to the field of plant disease diagnosis, highlighting the importance of future investigations in this area.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**REFERENCES**

Peng Jiang, Yuehan Chen, Bin Liu, Dongjian He, Chunquan Liang,’ Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks’, (Volume: 7), pp. 06 May 2019

Zhou, R., Kaneko, S., Tanaka, F., Kayamori, M., Shimizu, M., ‘Disease detection of Cercospora Leaf Spot in sugar beet by robust template matching’, Computers and Electronics in Agriculture, Volume 108, pp.58-70, 2014

Barbedo, J.G.A., Godoy, C.V., ‘Automatic Classification of Soybean Diseases Based on Digital Images of Leaf Symptoms’, SBI AGRO, 2015

Bashish, D.A., Braik, M., Ahmad, S.B., ‘A Framework for Detection and Classification of Plant Leaf and Stem Diseases’, International Conference on Signal and Image Processing, pp. 113-118, 2010

Punajari, J.D., Yakkundimath, R., Byadgi, A.S., ‘Image Processing Based Detection of Fungal Diseases in Plants’, International Conference on Information and Communication Technologies, Volume 46, pp. 1802-1808, 2015

Zulkifli Bin Husin, Abdul Hallis Bin Abdul Aziz, Ali Yeon Bin Md Shakaff Rohani Binti S Mohamed Farook, “Feasibility Study on Plant Chili Disease Detection Using Image Processing Techniques”, 2012 Third International Conference on Intelligent Systems Modelling and Simulation

Chunxia Zhang, Xiuqing Wang, Xudong Li, “Design of Monitoring and Control Plant Disease System Based on DSPFPGA”, 2010 Second International Conference on Net- works Security, Wireless Communications and Trusted Computing.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_