Image Classification Using Machine Learning and Deep Learning

Report submitted in fulfillment of the requirements

for the Exploratory Project of

Second Year B.Tech.

by

DHEERAJ YADAV

21075026

GAURAV YADAV

21075034

RUCHIRA NASKAR

21075072

Under the guidance of

DR. RAJEEV SRIVASTAVA

Department of Computer Science and Engineering

INDIAN INSTITUTE OF TECHNOLOGY (BHU) VARANASI

Varanasi 221005, Uttar Pradesh, India

May 2023

DECLARATION

We, the undersigned students, hereby declare that the project entitled "**Image Classification Using Machine Learning and Deep Learning**" submitted by us to the Indian Institute of Technology (BHU) Varanasi during the academic year 2022-23 in fulfillment of the requirements of the Exploratory Project for the award of Degree of Bachelor of Technology in Computer Science and Engineering is a record of bonafide project work carried out by us under the guidance and supervision of Dr. Rajeev Srivastava.

We have worked collaboratively throughout the project and followed the ethical guidelines for conducting research and ensured that our methods and results were accurate and reliable. We have also maintained a detailed record of our research methodology, data collection, and analysis procedures, and given due credit to external sources through citations.

We further declare that the work reported in this project has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other University.

Dheeraj Yadav (21075026)
Gaurav Yadav (21075034)
Ruchira Naskar (21075072)

Place: Varanasi, India
Date: May 4, 2023

Department of Computer Science and Engineering

INDIAN INSTITUTE OF TECHNOLOGY (BHU) VARANASI

Varanasi 221005, Uttar Pradesh, India

CERTIFICATE

This is to certify that the report entitled "**Image Classification Using Machine Learning and Deep Learning**" submitted by **Dheeraj Yadav (21075026)**, **Gaurav Yadav (21075034)**, and **Ruchira Naskar (21075072)**, carried out in the Department of Computer Science and Engineering, Indian Institute of Technology (BHU) Varanasi is a bonafide record of the project work carried out by him under our guidance and supervision.

Dr. Rajeev Srivastava

Department of Computer Science and Engineering,

Indian Institute of Technology (BHU) Varanasi,

Varanasi 221005, India

Place: Varanasi, India

Date: May 4, 2023

CONTENTS

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to Dr. Rajeev Srivastava and Dr. SK Singh, Head of the Department of Computer Science and Engineering, for their guidance and Sir Rajat Arya for his support.

Dheeraj Yadav

Gaurav Yadav

Ruchira Naskar

Place: Varanasi, India

Date: May 4, 2023

ABSTRACT

The application of machine learning and deep learning concepts to devise solutions to various real-life problems is becoming increasingly necessary due to the growing volume and complexity of related datasets. There have been numerous developments in the field of computer vision and image processing which have helped in the simplification of complex problems related to different spheres of the world. Utilizing an appropriate and applicable model individually or in combination with other training models, systems have been built to work on bringing out and improving the outcome of various data-based problems. We have applied combinations of Convolutional Neural Networks (CNNs), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and Random Forests to classify foliar diseases in apple trees, Bidirectional Long Short-Term Memory (Bi-LSTM) and CNN for violence detection in smart surveillance, and Principal Component Analysis for classification of cancer cells.

CHAPTER 1

INTRODUCTION

1.1 Overview

Deep learning and machine learning techniques have transformed the field of computer vision and image classification. These technologies have made it possible to develop systems that can interpret and analyze digital images and videos with high accuracy, which has opened doors to numerous opportunities across a variety of industries and also helped solve various pressing problems of importance. Machine Learning algorithms can build solutions to various tiresome data related to daily-life importance and vital spaces like medicine, education, and welfare roles. For example, the detection of cancer cells and their classification holds immense importance in the field of medicine and research, or detection and recognition systems to make concerning systems more convenient.

A major application of deep learning and machine learning in computer vision and image classification is object recognition and detection. Deep learning algorithms are implemented to identify specific features and patterns in images and videos, enabling them to classify objects in an accurate manner. This technology is widely used in medical needs, botanical requirements, self-driving cars, surveillance systems, industrial automation systems, and others.

A neural network is widely used in deep learning. It is a type of artificial intelligence that resembles the structure and function of the human brain. Neural networks are a powerful tool for solving complex problems and have found innumerable applications in various fields, and their popularity is increasing exponentially. Neural networks are capable of learning to recognize complex patterns in data and make accurate predictions or classifications based on that data. They are highly useful in cases where traditional machine learning algorithms struggle, like with image and speech recognition. On being provided with a large dataset of labeled examples, a neural network can learn to recognize patterns in that data and use that knowledge to make predictions or classifications on new, and unseen data. Neural networks can also automatically learn to extract useful features from data, reducing the need for manual feature engineering needed for making predictions. The ability of neural networks to solve complex problems has made them a valuable tool in the field of computer science.

1.1.1 Classification of Cancer Cells

For an accurate diagnosis, prognosis, and proper treatment planning, cancer cells must be classified. Cancer is a complicated illness with numerous subtypes, each of which has particular molecular and genetic characteristics. It is essential to categorize cancer cells into various subtypes in order to predict their clinical behavior, choose the best course of treatment, and track the development of the disease. Principal Component Analysis (PCA), a machine learning technique, can assist in the classification of cancer cells by lowering the dimensionality of the data and locating the most pertinent features.

1.1.2 Classification of Foliar Diseases in Apple Trees

Digital imaging and machine learning have recently demonstrated great promise for accelerating the diagnosis of plant diseases. Much of the world now has ready access to a smartphone with an integrated digital camera that can be used to take high-quality pictures of disease symptoms. The digital imaging revolution has already greatly expanded opportunities in many spheres of social and professional life. To classify diseases based on digital images of symptoms, computer vision techniques are being developed. These techniques look for connections and visual patterns for classification and grouping employing human expertise and machine learning algorithms [5]. In our venture, we have utilized Random Forests, XGBoost, Densenet CNN, CNN, and SVM in different combinations in attempts to research on the developed methods to classify the foliar diseases and look for more accuracy.

1.1.3 Violence Detection in Smart Surveillance

The adoption of artificial intelligence and neural networks-based learning has made human lives more secure with an increase in the number of installed security cameras around the world adapted to smart surveillance objectives. These cameras are used to monitor people's activities, detect objects, safeguard people's assets, and determine the circumstances surrounding specific actions using CCTV footage. The detection and recognition of ongoing unhealthy events lower crime rates and improve the security and safety of the environment [6]. We have attempted to develop our work in this domain with the help of Convolutional Neural Networks and BiLSTM.

1.2 Motivation

Over the past two decades, machine learning has advanced significantly, from a curious idea in-lab tool to a useful tool with widespread commercial application. Machine learning has become the approach of choice in artificial intelligence (AI) for creating useful software for computer vision, speech recognition, natural language processing, robot control, and other applications [2]. The recent surge of interest in gaining knowledge about deep learning methods lies in the fact that they have been shown to outperform previous previously existing technologies in various tasks, as well as utilizing and learning on the abundance of complex data from different sources (e.g., visual, audio, medical, social, and sensor) [7].

The primary motivation for taking on this project was our desire to venture into an exciting field of research and space of immense futuristic capacity. The chance to learn more about Computer Vision and Image Classification we had not been exposed to earlier pushed us to work harder. Studying the effects of implementations of various machine learning and deep learning algorithms to solve real-life problems across domains has helped us to delve deeper into the impacts of our work in practicality and motivated us to make them better.

CHAPTER 2

LITERATURE REVIEW

The objective of a literature review is to gather current, pertinent research on the subject of our choice and to synthesize it into a comprehensive overview of the existing knowledge in the area. This, therefore, equips us to present our own case or carry out an independent study on the subject. We have, thus, read and analyzed the various processes and methodologies enlisted across research-related papers based on machine learning and deep learning, and obtained knowledge and resources to assist in our findings.

2.1 Classification of Cancer Cells

Early detection and diagnosis of cancer cells is crucial since it considerably improves the patient's prognosis and enhances the likelihood that the disease will be successfully treated. Cancer is a condition that develops when body cells begin to multiply and grow uncontrollably, creating a lump or tumor. Cancer cells have the ability to proliferate and spread throughout the body if unrecognized and can possibly harm organs and tissues seriously if left untreated.

In 2020, Amrane et al. proposed Naive Bayes (NB) classifier and K nearest neighbor (KNN) algorithms to be used for breast cancer classification. The most common classification for breast cancer is binary (benign cancer/malign cancer), which helps pathologists discover a systematic and objective prognosis. Nowadays, machine learning (ML) techniques are widely used in the problem of breast cancer categorization. A comparison between the two new implementations was proposed, and their accuracy was evaluated using cross-validation. Results showed that KNN achieved the highest accuracy of 97.51% with the lowest error rate, followed by the NB classifier (96.19%).

Principal Component Analysis and Logistic Regression on the UCI, Breast Cancer Wisconsin (Diagnostic) Dataset (2020) has yielded best achievement of 97% on 30% validation dataset (from features based on perimeter and concavity).

2.2 Classification of Foliar Diseases in Apple Trees

Apple trees with foliar disease suffer significant health and productivity losses. Accurate disease diagnosis is essential for effective early treatment and avoidance of yield loss. A foliar disease identification model can help farmers tell the difference between healthy apple leaves and leaves that have apple rust, apple scab, or other diseases [3]. Various approaches have been proposed for the classification of foliar diseases in apple trees, including traditional morphological methods, molecular methods, spectroscopy, and machine learning algorithms.

In 2019, Jiang et al. proposed a deep learning-based method for identifying and classifying apple leaf diseases using improved CNNs. A large dataset of images of apple leaves infected with different diseases was used to train the CNNs to classify them into different categories of diseases. The proposed method achieved high accuracy rates (mean average accuracy of 78.80%) in the classification of apple leaf diseases.

In 2021, Bansal et al. proposed an ensemble of pre-trained DenseNet121, EfficientNetB7, and EfficientNet NoisyStudent, with the aim to classify leaves of apple trees into either healthy, apple scab, apple cedar rust, or multiple diseases, with the help of its images. The proposed model achieved an accuracy of 96.25% on the validation dataset.

In 2022, Yadav et al. proposed an AFD-Net model for leaf disease classification in apple trees and the results turned out to be the highest values of 98.7% accuracy for Plant Pathology 2020 and 92.6% for Plant Pathology 2021 compared to other deep learning models in the original and extended datasets.

2.3 Violence Detection in Smart Surveillance

The demand for a system that can automatically detect violence and suspicious activity has increased with the use of surveillance cameras to monitor human activity. The study of abnormal and violent action detection has become an active area of study in computer vision and image processing for new researchers [4].

In 2007, Shah et al. proposed a deep learning-based approach for violence detection in surveillance videos was proposed. The authors applied background subtraction and tracking algorithm on a dataset of videos containing violent and non-violent events and achieved around 88% accuracy rates in detecting violent events.

In 2019, Song et al. proposed a novel violent video detection scheme based on the modified 3D ConvNet was proposed. The scheme obtained competitive results like 99.62% on hockey fights, 99.97% on movies, and 94.3% on crowd violence.

CHAPTER 3

APPROACHES USED

3.1 Classification of Cancer Cells

3.1.1 Principal Component Analysis and Logistic Regression

An unsupervised learning algorithm called principal component analysis (PCA) is used in machine learning to reduce dimensionality. With the aid of orthogonal transformation, it is a statistical process that transforms the observations of correlated features into a set of linearly uncorrelated features. These new transformed features are called the Principal Components. PCA is one of the widely used tools for exploratory data analysis and predictive modelling. It is a method for identifying significant patterns in the provided dataset by lowering the variances. Typically, PCA looks for the surface with the lowest dimensionality to project the high-dimensional data. PCA functions by taking into account each attribute's variance because a high attribute demonstrates a good split between classes, which lowers the dimensionality. Image processing, movie recommendation systems, and power allocation optimization in various communication channels are some examples of practical applications of PCA. Since it uses a feature extraction technique, it keeps the crucial variables and discards the unimportant ones.

Logistic Regression is a widely used Machine Learning algorithm, under the category of Supervised Learning. Using a predetermined set of independent variables, it is used to predict the categorical dependent variable. As a result, the outcome must have a discrete or categorical value. Rather than providing the exact values of 0 and 1, it provides the probabilistic values that fall between 0 and 1. It can be either Yes or No, 0 or 1, true or false, etc. With the exception of how they are applied, logistic regression and linear regression are very similar. While logistic regression is used to solve classification problems, linear regression is used to solve regression problems. In logistic regression, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1), rather than a regression line. The logistic function's curve shows the likelihood of various things, including whether or not the cells are cancerous, whether or not a mouse is obese based on its weight, etc. As it can provide probabilities and classify new data using both continuous and discrete datasets, logistic regression is a significant machine learning algorithm. Logistic regression can be used to classify the observations using different types of data and easily identify the variables that will classify the best.

Once scaling is done, the PCA model is fitted, and the features are transformed into the PCs. As there are 30 features, there can be 30 PCs at maximum. The first two are taken into consideration. The plot obtained can be used to build intuition for the predictive capability of the given data. Here, it is seen that the entire dataset will allow us to separate tumors on the basis of them being malignant and benign tumors. There might still be some outliers as only the first two PCs have been taken into consideration. A model trained on the full feature set might be able to produce a better outcome.

PCA can also be used to compare different groups of features. For example, let two groups be there. Here, Group 1 has all the features based on cell symmetry and smoothness features, and Group 2 has all the features based on perimeter and concavity. PCA can be used to gain an intuition about which group is more adept at making predictions. PCA is conducted on each group separately. This will give us two sets of PCs and we select PC1 and PC2 to represent each of these feature groups. The group with more distinct clusters defines the features in it to be better at making predictions.

In the project, on 70% training data and 30% test data, the accuracy on the test set for Group 1 with more overlap was 74%, and in comparison, the accuracy for Group 2 with less overlap was 97%. Thus, the features in Group 2 are better at predicting the PCA results. PCA can be used to gain a better view of the data before modelling is done. It gives an idea of the approximate classification accuracy. It can also be used to determine which features are predictive, which helps in feature selection.

3.1.2 Principal Component Analysis and Linear Discriminant Analysis

Principal Component Analysis is a popular unsupervised learning method for lowering the dimensionality of data. While minimising information loss, it also improves interpretability. It makes data easy to plot in 2D and 3D and aids in identifying the dataset's most important features. PCA helps in finding a series of linear combinations of variables.

The most popular dimensionality reduction method in supervised learning is linear discriminant analysis. In essence, it serves as a preprocessing step for applications like pattern classification and machine learning. It reduces overfitting and computational costs by projecting the dataset into a moderately dimensional space with a genuine class of separable features.

The LinearDiscriminantAnalysis class can be imported from the sklearn.discriminant_analysis module. An object of the LinearDiscriminantAnalysis class is initialized with n_components set to 1, indicating that we want to reduce the dimensionality of the data to 1 dimension. The training and test data are transformed into 1-dimensional space using the fit_transform() and transform() methods respectively. This method is called Linear Discriminant Analysis (LDA) which is a technique for dimensionality reduction that is commonly used in classification problems to improve model accuracy. It finds a linear combination of features that characterizes or separates two or more classes of objects or events.

3.2 Classification of Foliar Diseases in Apple Trees

Model Architecture

We used a deep learning model based on the DenseNet121 architecture pre-trained on the ImageNet dataset. The model is designed for a new task and has four output classes.The input shape is set to 100x100 pixels with three color channels.We set all layers in the DenseNet121 model to be trainable.

Then, two pooling layers are added to the output of the model: a global average pooling layer and a global max pooling layer. The two output tensors from these layers are combined by taking their element-wise minimum. This combination of pooling layers helps to extract both local and global features from the input image.

Then we applied dropout regularization with a drop factor of 0.5 to the output tensor. Then, three dense layers are added with different activation functions (sigmoid and relu) and different numbers of neurons. Each of these dense layers is followed by a dropout layer with a drop factor of 0.3 to prevent overfitting.

Finally, the output layer is defined as a dense layer with four neurons and a softmax activation function. The inputs and outputs of the model are defined, and the model is compiled using the Adam optimizer with a learning rate of 0.001 and categorical cross-entropy loss

Then we used four different types of classifiers and compared their accuracies. These classifiers are:

1.DenseNet CNN

2.CNN + XGBOOST

3.CNN + Random Forest

4.CNN + SVM

3.2.1 DenseNet CNN

One of the most recent advancements in neural networks for visual object recognition is DenseNet. ResNet and DenseNet are quite similar, but there are some key distinctions. While DenseNet concatenates (.) the output of the previous layer with the output of the subsequent layer, ResNet uses an additive method (+) that merges the previous layer (identity) with the subsequent layer. DenseNet was developed especially to improve the degrading accuracy caused by the vanishing gradient in high-level neural networks.

The model begins by taking the DenseNet121 model pre-trained on ImageNet as a base and adding global average and max pooling layers to obtain a merged output. The merged output is passed through dropout layers to reduce overfitting and fed into a series of dense layers with varying activation functions and dropout rates. The final dense layer has 4 units with a softmax activation function to produce the output probabilities for the 4 classes of apple foliar disease. The model is compiled using the Adam optimizer with a learning rate of 0.001 and evaluated using the categorical cross-entropy loss function, accuracy, precision, and recall metrics.

3.2.2 Convolutional Neural Networks and Extreme Gradient Boosting

A neural network type called a convolutional neural network, or CNN or ConvNet, is particularly adept at processing data with a grid-like topology, like an image. A digital image is a binary representation of visual data. It is made up of a grid-like arrangement of pixels, each of which has a pixel value to indicate brightness and what color it should be. The moment we see an image, the human brain begins processing a massive amount of data. Every neuron has a distinct receptive field and is connected to other neurons so that they collectively cover the entire visual field. Each neuron in a CNN processes data only in its receptive field, similar to how each neuron in the biological vision system responds to stimuli only in the constrained area of the visual field known as the receptive field. Lines, curves, and other simpler patterns are detected first by the layers, followed by more intricate patterns like faces and objects. One can enable sight to computers by using a CNN, thus helping in computer vision.

XGBoost, which stands for Extreme Gradient Boosting, is a scalable, distributed gradient-boosted decision tree (GBDT) machine learning library that provides parallel tree boosting and leads machine learning library for regression, classification, and ranking-related problems.

The XGBoost classifier model, with a maximum depth of 50 and 500 estimators, is initialized with hyperparameters, such as maximum depth and number of estimators. The fit() method is called to train the model on the given training data, where x_tr contains the input features and y_train contains the ground truth labels that are one-hot encoded. The argmax function is used to convert the one-hot encoded labels to categorical labels. During training, XGBoost builds decision trees sequentially, attempting to correct the mistakes of the previous trees, while using gradient boosting to minimize the loss function. The classifier's objective is to optimize the margin between the decision boundary and the points of each class, thus achieving good separation of the classes. Once trained, the model can be used to classify new, unseen data, making predictions based on the learned decision boundaries.

3.2.3 Convolutional Neural Networks and Random Forest

A deep learning neural network designed for processing structured arrays of data, such as portrayals, is known as a convolutional neural network, or CNN.

Random Forest is a popular and frequently applied algorithm by data scientists. Supervised machine learning algorithms like random forest are frequently employed in classification and regression issues. On various samples, it constructs decision trees and uses their average for classification and majority vote for regression. The Random Forest Algorithm's ability to handle data sets with continuous variables, as in regression, and categorical variables, as in classification, is one of its most crucial features. It performs better for classification and regression tasks.

3.2.4 Convolutional Neural Networks and Support Vector Machine

Deep Learning has established itself as a pivotal tool over the last few decades due to its capacity for handling huge amounts of data. Particularly for pattern recognition, hidden layer technology is more popular than conventional methods. Convolutional Neural Networks, often known as CNN or ConvNet, are among the most widely used deep neural networks, mainly for computer vision applications.

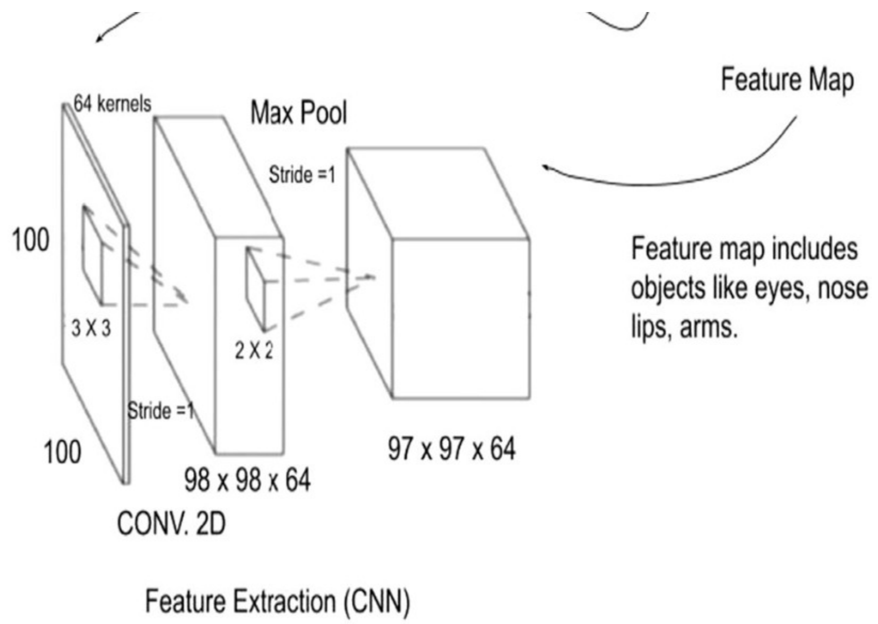
Support Vector Machine, or SVM, is used to solve Classification and Regression problems. However, it is primarily employed in Machine Learning Classification issues. The SVM algorithm's objective is to establish the best line or decision boundary that can divide n-dimensional space into classes, allowing us to quickly classify fresh data points in the future. A hyperplane is the best decision boundary.

SVM (Support Vector Machine) algorithm is used to create a model that can classify data into multiple categories. The 'decision_function_shape' parameter is set to 'ovo' which means it will use the One-vs-One approach for multi-class classification. Then, the model is trained on the training data (x_tr and y_train) and can be used to predict the class of new data.

3.3 Violence Detection in Smart Surveillance

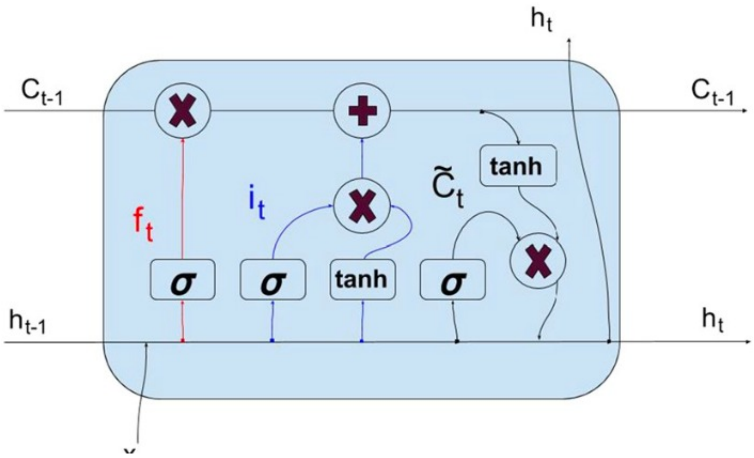
3.3.1 Convolutional Neural Networks and Bidirectional Long Short-Term Memory

A neural network type called a convolutional neural network, or CNN or ConvNet, is particularly adept at processing data with a grid-like topology, like an image. Each neuron in a CNN processes data only in its receptive field, similar to how each neuron in the biological vision system responds to stimuli only in the constrained area of the visual field known as the receptive field. Lines, curves, and other simpler patterns are detected first by the layers, followed by more intricate patterns like faces and objects. One can enable sight to computers by using a CNN, thus helping in computer vision.



General CNN Structure

A bidirectional LSTM, also known as a biLSTM, is a sequence processing model that consists of two LSTMs, one of which receives input forward and the other of which receives it backward. With the help of BiLSTMs, the network has access to more information, which enhances the context that the algorithm has (for instance, by letting it know what words come before and after a word in a sentence).



Basic LSTM Cell

Data Preprocessing

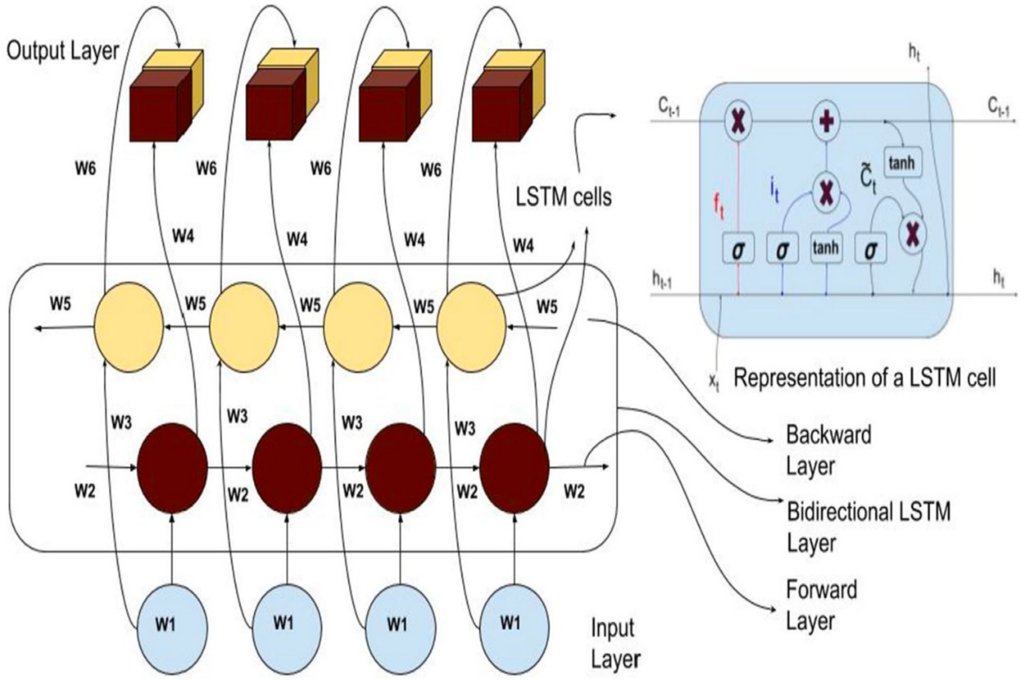
The extracted frames from the videos are reconfigured to a size of 100 x 100 pixels (referred to as $(x \times y)$). Each row of the Numpy3 array containing the training data corresponds to a video sequence or pattern. A series of movements and actions could be included, such as whether an arm movement is a punch or a handshake, etc. To extract a sequence, two frames are necessary as a minimum [1]. To extract the temporal features, we used 10 consecutive frames (referred to as n). The total sample count (denoted by N) is equal to the number of these sequences that are present in the dataset ((total frame count)/(number of frames to be taken into account in a sequence)) NumPy allows an arbitrary value of 1 to be used for simplicity. A structure containing a sequence of 10 consecutive frames with their respective class labels is prepared. The shape of the training data is $(-1, N, x, y, c)^4$. Here, c represents the number of channels in each frame.

Training Methodology

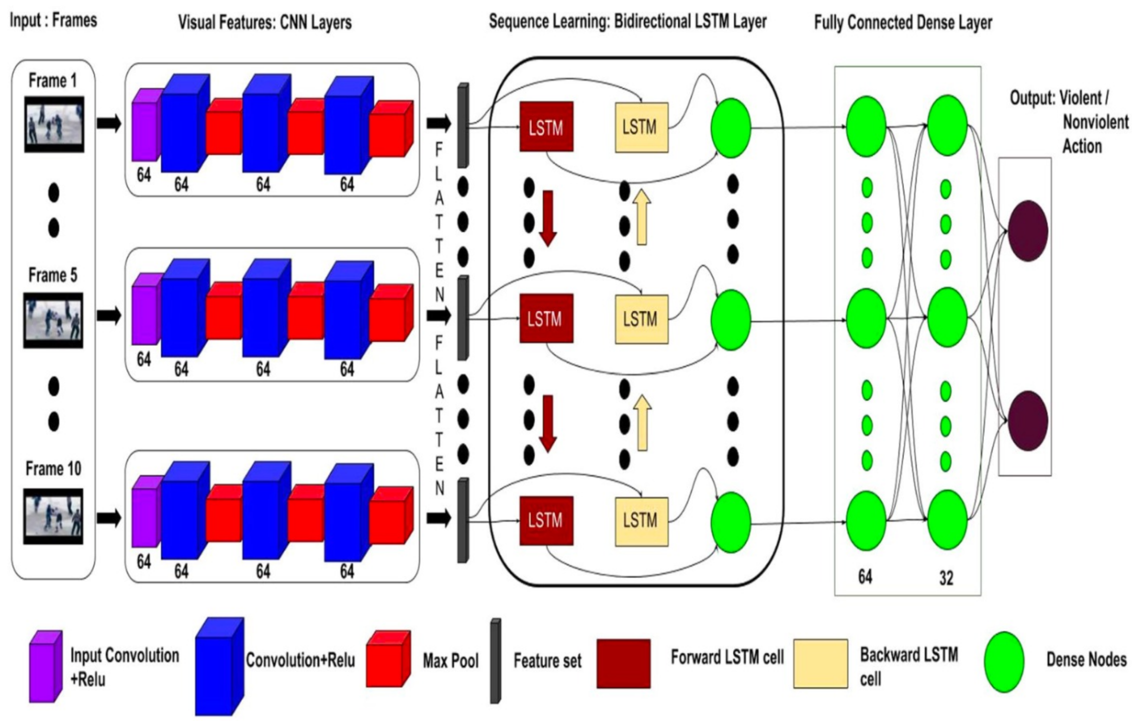
To extract the spatial and temporal features, a set of 10 consecutive frames with a dimension of $(100 \times 100))$ were sent to the model. A stochastic gradient descent optimizer has been used, with a learning rate of 0.01 and a decay of $1e^(- 6)$. This study uses the sparse categorical cross-entropy loss function. Instead of one-hot encoding, we have used 0 or 1 as class labels in this multi-class classification problem, with a batch size of 5 samples at a time. The datasets are split into training and testing portions in a 9:1 ratio. For 25 epochs, the entire model was built from scratch and trained in order to maintain its low computational cost.

Model Architecture

The proposed model architecture for violence detection in surveillance videos consists of three sub-parts: a Convolutional Neural Network (CNN), a Bidirectional Long Short-Term Memory (BiLSTM) layer, and Dense layers. The CNN includes an input convolutional layer, followed by three layers of convolution and max pooling. The kernel size for each convolutional layer is 3×3, and 64 kernels are used in each layer. The output from each convolutional layer after passing through the ”relu” activation function is max pooled to extract the features. The BiLSTM layer processes temporal features in both forward and reverse directions to detect sequences in consecutive frames. The LSTM cells are used to reconsider a part of previously trained features, and the bidirectional mechanism adds robustness to the model. Finally, the Dense layers are used to add random weights and features to find the best accuracy over a certain number of epochs. Overall, the proposed model architecture can classify violent or non-violent actions in surveillance videos accurately while consuming less computational time.



Bidirectional LSTM Cell



Model Architecture

CHAPTER 4

DATASET

4.1 Classification of Cancer Cells

The Wisconsin Diagnostic Breast Cancer (WDBC) dataset is a well-known benchmark dataset in the field of machine learning that aims to perform cancer detection with higher accuracy. It was obtained by Dr. William H. Wolberg, a medical researcher at the University of Wisconsin Hospitals, and contains a total of 569 instances, each representing a biopsy of breast tissue. Each instance in the dataset is characterized by 32 attributes, including measures of the size, shape, and texture of the cell nuclei in the biopsy sample. The target variable in the dataset is the diagnosis of the biopsy, which can be either malignant (cancerous) or benign (non-cancerous).

The WDBC dataset has been widely used in research on breast cancer diagnosis and machine learning, as it provides a rich and diverse set of features that can be used to train and evaluate models for cancer detection. In particular, it has been used extensively to study the performance of various classification algorithms, including decision trees, support vector machines, and neural networks, for distinguishing between malignant and benign biopsies.

A benefit of the WDBC dataset is that it is relatively small and well-curated, making it convenient and less prone to overfitting. It is available for free download from the University of California, Irvine Machine Learning Repository, to make it accessible to researchers and practitioners all over the world and assist in their study of cancer cell detection.

The WDBC dataset has played a vital role in progressing our understanding of breast cancer diagnosis and the application of machine learning techniques to solve medical problems.

4.2 Classification of Foliar Diseases in Apple Trees

The Plant Pathology 2020 dataset on Kaggle is a comprehensive collection of labeled images of healthy and diseased apple leaves that were used to classify different foliar diseases in apple trees. The dataset consists of a total of 18,232 images captured from apple orchards in the United States and Canada over several years. The dataset was created keeping in mind the objectives of using the images of the training dataset to accurately classify a given image from the testing set into different diseased categories or a healthy leaf, to accurately differentiate between many diseases, possibly multiple ones on a single leaf, to deal with rare classes and novel symptoms, to address light, shade, depth perception—angle, physiological age of the leaf, and to incorporate expertise in the identification, annotation, quantification, and proper guidance of computer vision to look for relevant features during learning.

4.3 Violence Detection in Smart Surveillance

The standard Hockey Fights dataset containing clips from ice hockey games has been used in the project. 500 violent and 500 non-violent clips with an average runtime of one second each are found in the collection. The videos shared similar backgrounds and subjects.

CHAPTER 5

CHALLENGES

5.1 Classification of Cancer Cells

5.1.1 Principal Component Analysis

- PCA assumes that the relationship between variables is linear whereas there there may be complex, nonlinear relationships between features in the dataset that are not captured by PCA.

- PCA can be difficult to interpret in terms of the original features, making it tough to gain insights into the underlying biology of the breast cancer dataset.

- PCA does not consider the target variable, which means that it may not identify the most important features for distinguishing between benign and malignant breast tumors.

5.1.2 Logistic Regression

- Logistic regression models are limited in complexity, and they may not be able to capture complex, non-linear relationships between the input features and the output variable.

- Logistic regression models can be sensitive to outliers, which can significantly influence the estimated coefficients and lead to biased predictions.

- Logistic regression models may not perform well on imbalanced datasets, where one class is much more prevalent than the other, as it will tend to favor the majority class, leading to poor predictive performance for the minority class.

- Logistic regression models require careful selection of input features, as irrelevant or redundant features can reduce the predictive performance of the model.

5.1.3 Linear Discriminant Analysis

- LDA assumes that the data is normally distributed and that the variance of each class is the same which might not hold in the breast cancer dataset.

- The breast cancer dataset has a relatively balanced number of samples between the two classes (malignant and benign tumors). However, in case one class was significantly underrepresented, LDA may not perform well.

- LDA can identify the most discriminative features for differentiating between the two classes, however it may not be able to capture complex interactions between features.

5.2 Classification of Foliar Diseases in Apple Trees

5.2.1 DenseNet CNN

- There may be a class imbalance, which could make it more difficult for DenseNet CNN to correctly classify samples from the minority classes.

- The small size of dataset can make it challenging for DenseNet to learn complex features and generalize well to new datasets.

- DenseNet CNN is a deep neural network architecture, which can be computationally expensive to train and evaluate and can be challenging while working with limited computational resources.

- DenseNet, can be difficult to interpret in terms of decision-making. This could be a challenge in the context of plant pathology, where it can be important to understand which features are most important for identifying different plant diseases.

5.2.2 Convolutional Neural Networks and Extreme Gradient Boosting

- Depending on the distribution of samples across classes, there may be a class imbalance, which could make it more difficult for XGBoost to correctly classify samples from the minority classes.

- If the number of features in the plant pathology dataset is very large, longer training times and increased computational complexity can result.

- Finding the optimal values for the hyperparameters can be time-consuming and computationally expensive

- XGBoost has high predictive power, but it can be difficult to interpret how the model is making its predictions.

5.2.3 Convolutional Neural Networks and Random Forest

- CNNs and random forest algorithms can be computationally expensive to train and evaluate if a large number of features are used.

- The CNN + random forest model can be prone to overfitting, particularly if not regularized properly.

- The CNN + random forest model can be difficult to interpret in terms of how it is making decisions.

5.2.4 Convolutional Neural Networks and Support Vector Machine

- The CNN and SVM model can be prone to overfitting, particularly if the SVM classifier is not regularized properly.
- CNNs and SVMs can be computationally expensive to train and evaluate, specially if a large number of features are used.
- The CNN + SVM model can be difficult to interpret in terms of its decision-making.

5.3 Violence Detection in Smart Surveillance

5.3.1 Convolutional Neural Networks and Bidirectional Long Short-Term Memory

- Video data is complex and dynamic, with variations in lighting, background noise, and camera angles, making it challenging to detect violent behavior accurately.
- The CNN-BiLSTM model is a computationally expensive deep learning model.
- The CNN-BiLSTM model, like other deep learning models, can be difficult to interpret in terms of decision-making.

CHAPTER 6

RESULT AND COMPARISON

6.1 Classification of Cancer Cells

Sl.no.	Algorithm	Accuracy for validation set
0	PCA + Logistic Regression	0.97
1	PCA vs. LDA	0.9649 (LDA) 0.9035 (PCA)

6.1.1 Comparison between different models

LDA and PCA differ in their goals and how they achieve them, which can lead to differences in performance on classification tasks.

PCA is a technique used to reduce the dimensionality of a dataset by projecting the data onto a new set of orthogonal axes that capture the largest amount of variance in the data. The resulting principal components can be used to represent the data in a lower-dimensional space, while still retaining most of the original information. However, PCA does not take into account the class labels of the data, and thus it may not be optimal for classification tasks where the goal is to maximize the separation between classes.

In contrast, LDA is a supervised learning technique that takes into account the class labels of the data and aims to find a lower-dimensional representation of the data that maximizes the separation between classes. Specifically, LDA seeks to find a linear combination of features that maximizes the ratio of between-class variance to within-class variance. By doing so, it creates a new feature space where the samples from different classes are as far apart as possible.

This property of LDA makes it particularly useful for classification tasks where the goal is to distinguish between different classes. When applied to such tasks, LDA can often achieve better accuracy than PCA.

6.2 Classification of Foliar Diseases in Apple Trees

Sl.no.	Algorithm	Accuracy for same testing data
0	DenseNet CNN	0.990240
1	CNN + XGBoost	0.986173
2	CNN + Random Forest	0.987393
3	CNN + SVM	0.988613

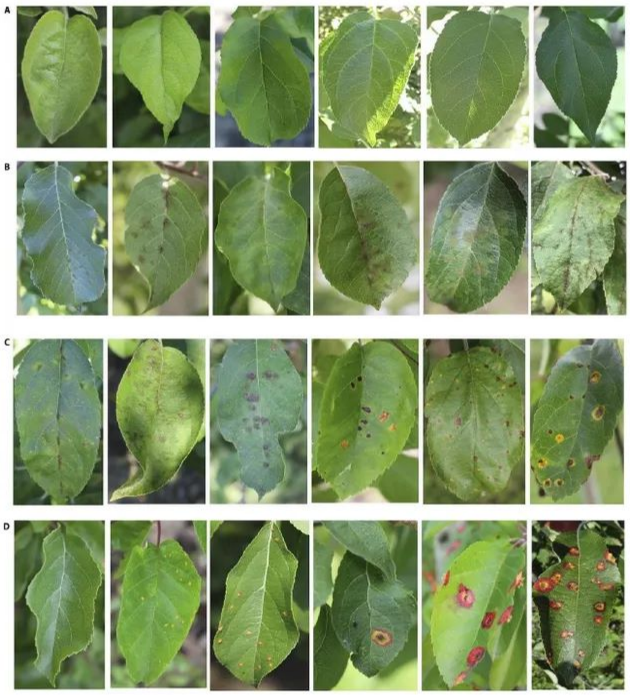
6.2.1 Comparison between different models

DenseNet is a deep neural network architecture that gave better accuracy than other models like SVM, XGBoost, and RandomForest because it can capture complex nonlinear relationships in the data that may be difficult for other models to capture.

DenseNet achieves this by using a dense connectivity pattern between layers, where each layer is connected to every other layer in a feed-forward fashion. This allows information to flow more easily through the network and enables the model to learn complex features at different levels of abstraction. In contrast, models like SVM, XGBoost, and RandomForest typically use a shallower architecture and rely on hand-engineered features to represent the input data.

Another advantage of DenseNet is that it can learn to reuse features across multiple layers, which reduces the number of parameters that need to be learned. This can be particularly beneficial when working with limited training data, as it allows the model to generalize better to new data.

RGB Channels for unhealthy part of leaf

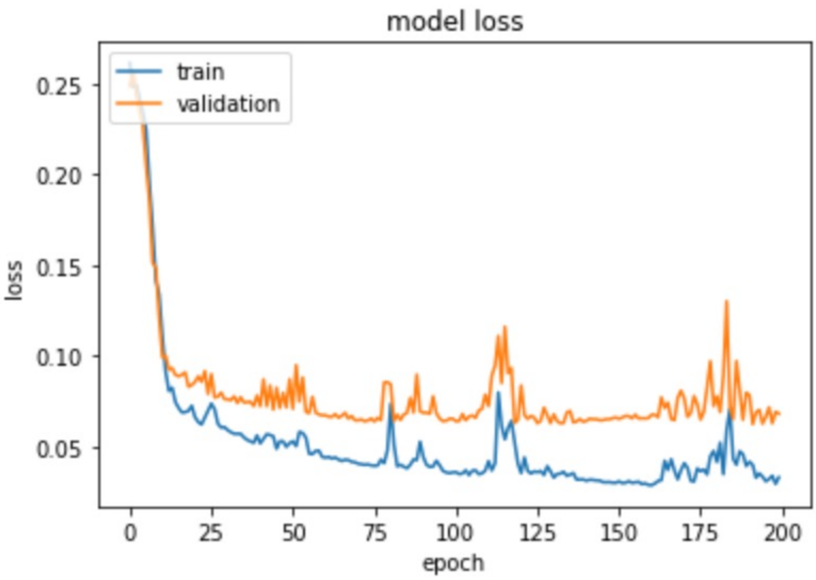


Sample images from the dataset showing healthy leaves (A), symptoms of apple rust (B), apple scab (C), multiple diseases on a single leaf (D)

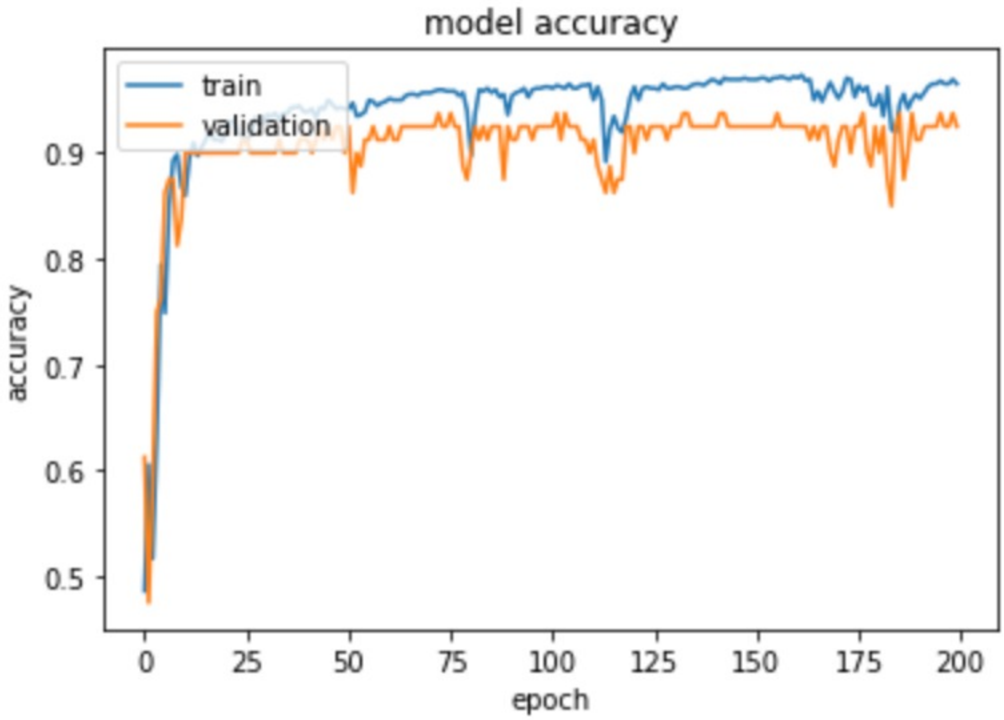
Plant Pathology 2020_ Healthy Images

6.3 Violence Detection in Smart Surveillance

Sl.no.	Algorithm	Accuracy for validation set
0	CNN + BiLSTM	0.9375



Plot of model loss vs epoch



Plot of model accuracy vs epoch

CHAPTER 7

CONCLUSION AND DISCUSSION

Using the Wisconsin Diagnostic Breast Cancer dataset, we have found out the advantages and limitations of both the PCA models with Logistic Regression and LDA, with PCA combined with Logistic Regression giving more accurate results on the validation set.

With the help of the Plant Pathology 2020 dataset, we have found out the effectiveness of Densenet CNN, along with the upsides and downsides of all the four models applied, including DenseNet CNN, CNN + XGBoost, CNN + Random Forest, CNN + SVM, in classifying foliar diseases in apple trees.

The deep learning model based on CNN and Bidirectional LSTM trained on the Hockey Fights dataset to detect violence in surveillance systems has shown an accuracy of 93.75%.

7.0.1 Future Directions

In the future, training and testing algorithms on larger and more varied datasets will improve the inclusivity and power of classification in the concerned solutions to real-life problems.

REFERENCES

[1] Rohit Halder and Rajdeep Chatterjee. Cnn-bilstm model for violence detection in smart surveillance. *SN Computer science*, 1(4):201, 2020.

[2] Michael I Jordan and Tom M Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.

[3] R Dhivya Praba, R Vennila, G Rohini, S Mithila, and K Kavitha. Foliar disease classification in apple trees. In *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, pages 1–5. IEEE, 2021.

[4] Muhammad Ramzan, Adnan Abid, Hikmat Ullah Khan, Shahid Mahmood Awan, Amina Ismail, Muzamil Ahmed, Mahwish Ilyas, and Ahsan Mahmood. A review on state-of-the-art violence detection techniques. *IEEE Access*, 7:107560–107575, 2019.

[5] Ranjita Thapa, Noah Snavely, Serge Belongie, and Awais Khan. The plant pathology 2020 challenge dataset to classify foliar disease of apples. *arXiv preprint arXiv:2004.11958*, 2020.

[6] Fath U Min Ullah, Mohammad S Obaidat, Amin Ullah, Khan Muhammad, Mohammad Hijji, and Sung Wook Baik. A comprehensive review on vision-based violence detection in surveillance videos. *ACM Computing Surveys*, 55(10):1–44, 2023.

[7] Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis, Eftychios Protopapadakis, et al. Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018, 2018.