International Journal of Scientific Research and Engineering Development-– Volume 7 Issue 1, Jan-Feb 2024

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

A Literature Review of Machine Learning Models for Corn Quality Classification and Regression

Mika Allyana M. Briones[1], Anajasnie Shamei F. Fung[2], Edwin R. Arboleda[3]

[1]Department of Computer and Electronics Engineering, Cavite State University, Indang, Cavite, Philippines
Email: mikaallyana.briones@cvsu.edu.ph

[2]Department of Computer and Electronics Engineering, Cavite State University, Indang, Cavite, Philippines
Email: anajasnieshamei.fung@cvsu.edu.ph

[3]Department of Computer and Electronics Engineering, Cavite State University, Indang, Cavite, Philippines
Email: edwin.r.arboleda@cvsu.edu.ph

--************************----------------------------------

Abstract:
Machine learning methods that utilize multiple models, known as ensemble methods, have proven to be superior to single models by accurately capturing the data distribution and minimizing prediction bias and instability. These methods have significantly transformed the agricultural sector, particularly in the area of corn quality classification and regression. This review aims to consolidate various machine learning applications in agriculture, specifically in assessing corn quality, by examining relevant databases and online resources. It includes a systematic review of articles from the past five years that utilized machine learning models such as Support Vector Machine (SVM), Random Forest, Linear Regression, Convolutional Neural Network (CNN), Artificial Neural Network (ANN), K-Means Clustering, Principal Component Analysis (PCA), and Decision Tree. The review concludes that the effectiveness of these models is contingent on the specific task, the nature of the data, and the unique requirements of the problem.

Keywords **—Machine Learning, Corn Quality, Corn Regression, Agriculture**

--************************----------------------------------

I. INTRODUCTION

The rapid global population growth increases food consumption, straining modern agriculture. This industry faces challenges like changing climate, depleting resources, shifting food habits, and health concerns. Adopting eco-friendly and efficient techniques is crucial. Precision agriculture, focusing on sustainability, optimal production, and environmental safety, has emerged as a solution. Technological advances, particularly in large dataset integration and analysis, have improved yield prediction capabilities. Machine learning has proven to be a faster and more adaptive prediction method compared to traditional crop modeling methodologies.[1],[2]

Machine learning ensembles, composed of groups of models, outperform individual models by accurately capturing data distribution and limiting prediction bias and volatility. AI, specifically machine learning, has significantly influenced agriculture by utilizing large and diverse data. Supervised learning trains machines using various models to make accurate predictions on new data. Unsupervised learning approaches enable machines to identify trends and connections in large data sets [3], [4].

A. Corn Quality Parameters

Businesses constantly introduce new or improved marketing strategies in an effort to achieve operational excellence in the present dynamic and static competitive landscape. Improving customer satisfaction is attained by consistently improving the characteristics of products or services with the goal of providing higher quality offers [5]. The word "quality" in the industry refers to a variety of characteristics that are thought to be connected to a company's financial performance metrics and general business success. Engaging with quality is considered a strategy for achieving higher returns [6]. On the other hand, Quality criteria offer an organized framework for assessing and preserving the quality of goods, services, or procedures. Their significance is not limited to, consumer delight; it also encompasses regulatory compliance, efficiency enhancements, and general corporate performance. With these quality parameters, it helps to achieve the necessary goal and guide the overall process needed to assess for an effective outcome.

Assessing the quality of corn is a critical aspect of agricultural and industrial practices. Corn quality measurement techniques are mostly based on the intended use of the grain. Various key metrics are measured to gauge the suitability of corn for diverse purposes, encompassing factors such as kernel size and uniformity, weight, maturity, color and appearance, moisture, starch content, etc. Various criteria are evaluated to determine the overall quality of corn production, and these characteristics are included in commonly measured corn quality metrics. These measures are essential for assessing corn's fitness for a variety of uses, such as wet milling, dry milling, industrial, animal feed, seed production, and food production, each of these applications has its own set of measuring methods that aid in selecting the most suitable corn for the intended use [7]. Numerous uses and varying end-user demand of corn make it challenging to establish defined quality criteria. The standards that are deemed necessary for human consumption could not coincide with those that are required for use in animal feed or industrial operations. The difficulty is in customizing the evaluation criteria to meet the unique needs of diverse applications while recognizing the unique attributes and features that are important to consumers, farmers, and the companies that process grain for a range of applications. Universal quality criteria for corn must be defined with flexibility and adaptability due to the inherent variety in end-user expectations. In general, corn kernel assessment evaluates constituent features (e.g., moisture, protein content, fiber, etc.) as well as visual features (e.g., impurity, shape, etc.). While visual characteristics are manually retrieved by skilled operators, constituent measurements are gathered with the use of equipment and machinery specifically designed for this purpose. This labor-intensive manual procedure cannot guarantee uniformity due to the variation in the operator's capacity for assessment [8]. These corn kernels can be classified into some of the following categories: good kernel, defective kernel, and impurity.

Traditional farming refers to agricultural techniques that have been used for centuries and are distinguished by physical labor, basic tools, and traditional knowledge. Natural and traditional operation of crop types are often used by a method in small scale subsistence farming which utilize low equipment [9]. It is evident that traditional farming is time-consuming, repetitive, labor-intensive and not accurate and efficient when it comes to measuring corn quality as it provides several issues, particularly for large volumes of corn. Also, disagreements in the subjectivity and consistency of the evaluators in the corn value chain may arise as it is prone to errors [10]. Its goal is to achieve consistent and reliable results from the evaluators however it is crucial to overcome these problems as each evaluators has different perspective and opinions. To address this challenges, a potential solution is to utilize computer vision and machine learning for automatic corn classification [11], it must be continuously improved through research and development of evaluating corn quality. Some researches shows a more accurate, efficient and non-destructive ways in agriculture that can be use in corn classification such as sensor technology, ML algorithms, and data management systems [12], [13]. However, difficulties in this method may continue based on the regional uniformity, data interpretation, and integration of technology in evaluating corn classification. Overcoming these obstacles is crucial.

B. General Application of Machine Learning in Agriculture

A range of machine learning applications have been identified in the field of agriculture. [14] categorizes these applications into plant monitoring, soil analysis, detection/prediction processes, and animal monitoring. [15] further expands on these, discussing crop and livestock management, water management, and soil management. Both studies highlight the potential for machine learning to enhance decision support and action in farming. [16] emphasizes the role of machine learning in disease diagnosis and prediction, while [17] focuses on yield and price prediction, as well as leaf disease detection.

Machine learning is increasingly being applied in agriculture to improve efficiency and productivity. [18] provides a comprehensive review of the techniques and applications of machine learning in agriculture, including artificial neural networks, support vector machines, and decision trees. These studies show the potential of machine learning in agricultural aspects.

Agriculture is crucial for a nation's prosperity, but challenges arise with growing populations, climate change, and resource scarcity. Precision agriculture, or smart farming, offers a new strategy to address these issues, with machine learning (ML) playing a key role. ML, combined with Internet of Things (IoT) connected agricultural machinery, is revolutionizing traditional farming practices and significantly improving outcomes. It facilitates real-time agricultural surveillance, predicts soil properties like moisture content and organic carbon levels, and analyzes real-time sensor data and historical patterns to predict crop yield. ML can also detect weeds and diseases in crops early, helping farmers take preventative action and reduce crop damage. By lowering the risks and costs associated with farming operations and improving decision-making efficiency and precision, ML is expected to significantly transform agricultural practices [19], [20], [21]

C. Relevance of Machine Learning in Corn Quality Prediction

Machine learning has been effectively used in predicting and improving corn quality and yield. It has shown high accuracy in classifying corn seed varieties using the MLP classifier and in predicting corn yield based on seeding date. The integration of topographic indices and remote sensing data with machine learning models has further enhanced the accuracy of corn yield prediction, demonstrating an understanding of yield's spatial variability. These studies collectively underscore the pivotal role of machine learning in enhancing the precision of corn quality prediction [22] and [23],[24], [25], [26], [27] and [28].

Machine learning (ML) has become significant in predicting corn quality, with applications that enhance the precision and productivity of agricultural practices. These applications include determining feature importance, using ensemble models, predicting yield, controlling quality, and monitoring in real-time. ML models have been highly successful in accurately predicting corn yields [29]-[30], By analyzing large amounts of data from various sources, detailed maps of crop growth, nutrient levels, and moisture content can be produced [31]. Machine learning models excel in quality control of corn production by predicting crucial product characteristics, surpassing traditional methods [32]. Studies have shown that ensemble models, which are combinations of several machine learning models, provide more accurate predictions than single models. These models enhance the representation of the underlying data distribution and reduce prediction bias and variance [29]. The integration of Machine Learning (ML) models with Internet of Things (IoT) technologies enables real-time monitoring of crop conditions, facilitating timely interventions and enhancing crop quality [30]. Quality control after harvest can be conducted using Machine Learning (ML) models. These models can forecast the quality of corn based on various factors, including moisture content, nutrient content, and the presence of pests or diseases [33].

D. Advantages and Limitations of Machine Learning Models in Agriculture

Machine learning (ML) models are increasingly used in agriculture due to their ability to enhance productivity, reduce costs, and improve decision-making. They are effective in precision farming, illness detection, and crop yield prediction. However, their effectiveness depends on the availability of large volumes of high-quality data, which can be challenging to collect in remote areas. Additionally, issues like class imbalance, data sparsity, and high dimensionality can complicate their application [34], [35], [36]. Despite these challenges, ML models have led to innovative methods that increase yield and provide 24/7 security for remote facilities [37]. However, the return on investment might not be immediate, the implementation might be costly, and not all agricultural settings have access to the necessary robust technological infrastructure [38]. Furthermore, ML models might not always adjust well to frequent environmental changes in the agricultural sector [39].

This review is based on the current understanding and application of ML in agriculture. As technology advances, some of these limitations may be addressed, and new advantages may emerge [31].

E. Historical Perspective on Corn Quality Evaluation

Corn being the most widely used maize, also known by various common names, including maize corn, Indian corn, sweet corn, and field corn. It goes by several names depending on how it is used, such corn on the cob, popcorn, and cornmeal. The varied nomenclature for this adaptable crop is influenced by regional and cultural differences. Corn is now the most important cereal crop in terms of worldwide output, having surpassed rice and wheat around a decade ago. The development of high-yielding conventional and genetically modified genotypes, as well as its greater tolerance to varied habitats, can be linked to corn's rise to dominance. Notably, maize has the highest grain output per hectare of any crop. Its significance extends far beyond just a human consumption staple; rather, it is of critical economic relevance on a worldwide basis. Corn is important not just as a human food source, but also as a component in animal feed formulations and as a raw material for a wide range of industrial goods and biofuels. Although corn is a staple crop in subsistence agriculture, developed countries that are seeing a simultaneous increase in demand for wheat flour and animal goods mostly use corn for animal feed [40], [41], [42].

Analyzing historical viewpoints on the evaluation of corn quality shows that evaluation techniques have changed significantly over time in response to advancements in science, technology, and agriculture. Early efforts have been often very basic, concentrating on the fundamental sensory assessment and visual examination. The need for more rigorous and consistent assessment techniques increased along with the global significance of corn. Even though they were fundamental, traditional manual approaches were prone to subjectivity and inconsistent assessments from assessors, which led to the investigation of technology interventions for increased precision and workflow [43]. The route of corn quality evaluation has been formed over time by obstacles, including the limited scope, time limits, and destructiveness of some approaches. Modern technologies, such as machine learning, have been included in quality evaluations in the last several decades, and this has greatly improved their accuracy and scope. This literature review aims to trace the historical trajectory of corn quality evaluation, shedding light on the evolution of methodologies and the persistent challenges that have informed current practices in the field.

F. Aims

This literature review aims to develop a comprehensive summary on the uses of machine learning techniques in the evaluation of corn quality.

This study covers the following topics, including possible gaps and issues occur in machine learning models and algorithm for classification and regression, viable solutions, assessment of corn quality characteristics, different sources of data and techniques used, and evaluation metrics. Also, it compares various machine learning models approaches in assessing the quality of corn. It is beneficial to students and farmers in the field of agriculture as it gives insights and provide latest knowledge in corn quality prediction.

II. METHODOLOGY

In the recent studies of machine learning about the classification and regression of various corn quality an extensive review was utilized in this literature paper. It takes notes the approaches, datasets, and model performance that utilized in different studies that might help for future study in this area with the possible findings and implications of this literature review.

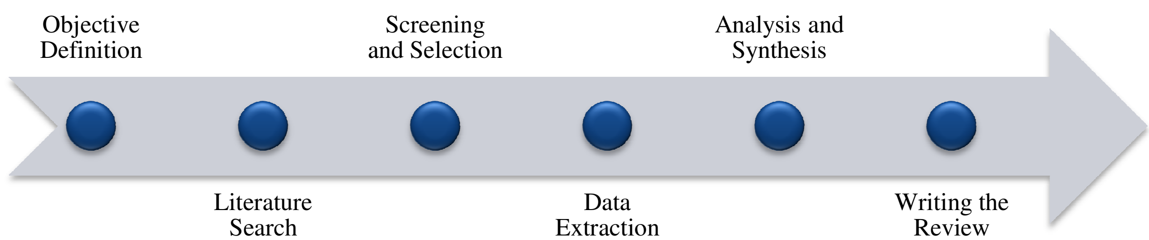


Fig. 1General System Diagram

This paper extensively collects and evaluate the previous studies that focuses on machine learning models in classification and regression of corn crop. We used online platforms and comprehensive database that is reliable to search various keywords to find related studies such as "machine learning," "corn quality," "classification," and "regression", however, studies that does not center on classification and regression of corn qualities using machine learning was regarded in this study since it does not associate with the focus of this study. Finding its common trends and gaps and pertinent data, it gives a comprehensive overview of machine learning when it comes to corn quality classification and regression. The collected data was organized in five sections to create this review, including introduction, discussion, machine learning models for corn quality classification and regression, results, and conclusion.

III. REVIEW OF SUPERVISED MACHINE LEARNING MODELS FOR CORN QUALITY CLASSIFICATION

A. Support Vector Machines

Support Vector Machine (SVM), a supervised machine learning algorithm that can be utilize in classification and regression of corn crop. When dividing a dataset into two separate groups for binary classification tasks, it is especially helpful. Finding a hyperplane that clearly divides the data points in an N-dimensional space, where N is the number of features, is the SVM's goal. This hyperplane serves as a decision boundary, classifying data points based on their relative position to it. The dimensionality of the hyperplane is determined by the number of features. For example, a line represents the hyperplane for two features, while a two-dimensional plane represents the hyperplane for three features. However, when the number of features surpasses three, visualizing the hyperplane becomes more complex [44], [45], [46].

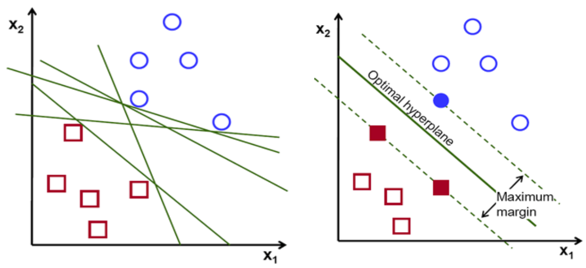


Fig. 2 Illustration of Support Vector Machine [46]

1) Ang Wu and colleagues developed a new method for classifying corn kernel quality using image analysis and a support vector machine in their study. Initially, both the support vector machine and back-propagation neural networks achieved a classification accuracy rate of 92.31% without parameter optimization. However, the accuracy improved with the use of optimization algorithms in the support vector machine. Both the support vector machine-genetic algorithm and support vector machine-particle swarm optimization achieved an average correct classification rate of 97.44%, while the support vector machine-grid search had a rate of 94.87%. The research found that the back-propagation neural network algorithm is outperformed by the support vector machine algorithm with parameter optimization, and that the grid search method is less effective for parameter optimization than the genetic algorithm and particle swarm optimization techniques [47].
2) The study [48] discusses the importance of nutrients like nitrogen, phosphorus, and potassium for corn plants. Various methods such as the Leaf Color Chart (LCC), Chlorophyll Meters Soil Plant Analysis Development (SPAD), and Soil Test Kit are used to examine these nutrients in corn leaves. The LCC method is preferred by farmers due to its lower cost. However, digital image processing, specifically the RGB extraction method of Hue, Saturation, Value (HSV), is proposed as a more efficient and cost-effective solution. The study uses the Support Vector Machine (SVM) for classifying the image results and achieves an accuracy of 80% in detecting nutrient content in corn leaves.
3) [49]proposes a classification method for assessing corn seed vitality using multisensor hyperspectral imaging. Hyperspectral images of waxy corn seeds were collected, and various preprocessing techniques were used to suppress noise in the raw spectra. Feature wavelengths were selected using principal component analysis, 2nd derivatization, and the successive projection algorithm. An SVM model was established and showed optimal performance when preprocessed by multiplicative scatter correction, achieving a training accuracy of 100% and a testing accuracy of 97.9167%. This method provides a new approach for nondestructive crop detection using machine learning.
4) The study [50] uses machine vision and machine learning to develop a method for rapid detection and classification of maize seeds based on variety purity. A computer vision system was designed to recognize five varieties of maize seeds, and an image processing algorithm was used to extract 16 important features from the seed images. Various machine learning algorithms were used to develop the classification model, with the SVM model achieving the highest accuracy. The results meet the needs of producers and consumers.

B. Random Forest

The Random Forest is a supervised learning algorithm that uses multiple decision trees for improved prediction accuracy. Each tree operates on different subsets of a dataset. The final output is chosen based on the majority of predictions from all trees. Each tree’s leaf nodes provide approximations of the probability distribution over the image classes, and each internal node contains a test for optimal data space partitioning. An image is classified by passing it through each tree and combining the resulting distributions [51], [52].

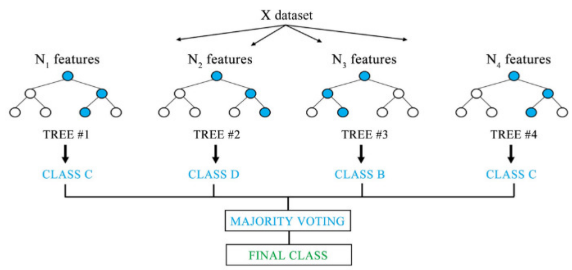
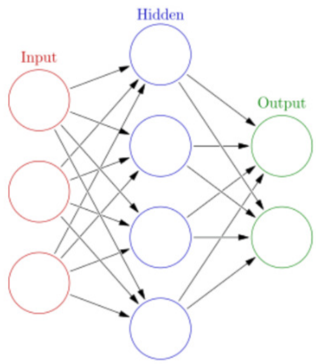


Fig. 3 Illustration of Random Forest [52]

1) Plant diseases significantly impact agricultural productivity. This paper [53] presents a method for early detection of crop diseases using AI techniques like Naive Bayes, Decision Tree, K-Nearest Neighbor, Support Vector Machine, and Random Forest. The methods were compared based on accuracy, and the Random Forest model was found to be the most accurate with an accuracy of 80.68%.
2) The study [54] proposes a method for improving crop yield prediction using the Geographically Weighted Random Forest Regression (GWRFR) approach. The GWRFR and five other machine learning algorithms were trained with different sets of features. The study found that the Geographically Weighted Random Forest Regression (GWRFR) with full-length features performed better than other algorithms.This method could potentially enhance yield predictions for various crops in different regions.
3) In their research, Junfeng Gao and peers uses a hyperspectral snapshot mosaic camera for weed and maize classification. A set of 185 spectral features was constructed and reduced using principal component analysis. A random forest classifier was developed using three different combinations of features. The optimal random forest model, with 30 important spectral features, achieved a mean correct classification rate of 1.0 for Zea mays, and lower rates for three types of weeds. This model outperformed the k-nearest neighbors model [55].
4) The study [56] uses Unmanned Aerial Vehicle (UAV) multispectral imagery to predict the canopy nitrogen weight of corn fields in south-west Ontario, Canada. Several machine learning models were tested, with the Random Forests model performing the best, achieving an R2 of 0.85 and an RMSE of 4.52 g/m2. The model was then used to produce maps showing the spatial variation of canopy nitrogen weight within each field at different dates.
5) Aeri Rachmad and colleagues' study develops a classification model using Random Forest for early-stage detection of corn diseases. The model uses fine and coarse features to capture various types of information for the classification process. The Local Binary Pattern method and Color Histogram are used in feature extraction. The model was tested on a dataset of 3,000 corn plant images and achieved an accuracy rate of 99.05% in identifying diseases [57].
6) The study [58] develops a new leaf image processing algorithm that uses Random Forest and leaf region rescaling to analyze nutrient and stress distributions across a corn leaf. This approach improves the quality of phenotyping measurements compared to traditional methods that average the spectrum across the whole canopy. The algorithm was tested on corn plants with different genotypes and nitrogen treatments, and it more clearly differentiated leaves from different treatments and genotypes. The algorithm could potentially improve the quality of other plant feature measurements.

C. Neural Network

Neural networks are machine learning techniques that use interconnected nodes or neurons



arranged in layers to simulate the structure of the human brain. This kind of deep learning procedure is based on the way neurons are arranged in the brain. The three primary layers of a neural network are the input layer, which receives data, the hidden layers, which carry out intricate calculations, and the output layer, which generates the outcome[59], [60], [61].

Fig. 4 Illustration of Neural Network [60]

1) The study by Zhang, Dai, and Cheng used hyperspectral imaging and a Deep Convolutional Neural Network (DCNN) to classify four corn seed varieties. The DCNN model outperformed other models, achieving 100% training accuracy, 94.4% testing accuracy, and 93.3% validation accuracy. The study suggests that DCNN can be effectively used for spectral data analysis and corn seed variety classification [62].
2) Abdul Waheed, Muskan Goyal, Ashish Khanna, Deepak Gupta, Aboul Ella Hassanien, and Hari Mohan Pandey proposes an optimized DenseNet model for recognizing and classifying corn leaf diseases. The DenseNet model, which utilizes fewer parameters than other Convolutional Neural Network (CNN) models, attained an accuracy of 98.06%. Despite using fewer parameters and less computation time, its performance is on par with established CNN architectures. This deep learning approach could potentially enable early disease detection, thereby enhancing crop health and increasing yield [63].
3) Linbai Wang, Jingyan Liu, Jun Zhang, Jing Wang, and Xiaofei develop a method for detecting defects in corn seeds using a watershed algorithm and a two-pathway convolutional neural network (CNN) model. RGB and near-infrared images of the seeds were used to train the model, which achieved an average accuracy of 95.63%, a recall rate of 95.29%, and an F1 score of 95.46%. The method outperforms the traditional one-pathway CNN with 3-channel RGB images and could be an effective tool for high-throughput quality control of corn seeds [64].
4) The study [65] uses a convolutional neural network to classify corn kernels with high accuracy. The Convex Hull method is added to increase focus on the convolution process by removing background images. The model uses a 34-layer architecture and dropout layers to save computation time. Data augmentation techniques are used to prevent overfitting. The model achieved an average accuracy of 99.33%, precision of 99.33%, recall of 99.33%, and F-1 score of 99.36%. The training time was 2 minutes 30 seconds. The use of the proposal area improved accuracy by about 0.3%.
5) The research paper [66] indicates a method precision farming technologies, specifically convolutional neural networks (CNN) and boosting techniques, to detect diseases and pests in corn crops. The aim is to classify disease manifestations with higher accuracy than existing methods. An ensemble of Adaptive Boosting cascaded with a decision tree-based classifier trained on features from CNN was used. This achieved an accuracy of 98% in classifying corn leaf images into four categories: Healthy, Common Rust, Late Blight, and Leaf Spot, representing an 8% improvement compared to using CNN alone.
6) In their study report, C. Naseeb Singh, M. Pareek, V. K. Tewari, L. K. Dhruw, and H. Dayananda Singh proposes an artificial neural network-assisted image-processing method for real-time classification of broken and whole maize kernels. Images of seed samples were captured, and image-processing operations were performed to extract morphological features.

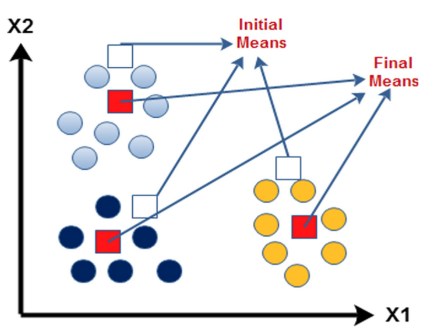
The ant colony optimization algorithm was used to select superior features. A multilayer perceptron neural network classifier was then used to identify the kernels, achieving a classification accuracy of 91.85% with an average processing time of 0.14 seconds per kernel.

7) The research paper [67] develops two artificial neural network (ANN) models to predict the quality parameters of corn used for ethanol production. The models consider various factors such as the type of corn hybrid, vegetation periods, agrotechnology levels, drying temperatures, and heating and pressure pre-treatments. The first model (ANN1) predicts the hectolitre weight, 1000-kernels weight, gelatinisation rate, and contents of glucose, reducing sugars, and ethanol. The second model (ANN2) predicts the corn weight and moisture. The models fit the experimental data well, with an overall r2 of 0.989 for corn kernel weight and moisture, and 0.856 for other outputs. The models can be used for multi-objective optimization in the corn kernel drying process.

IV. REVIEW OF UNSUPERVISED MACHINE LEARNING MODELS FOR CORN QUALITY CLASSIFICATION

A. K-Means Clustering

K-means clustering is a non-hierarchical method used in cluster analysis to group similar objects into clusters. Each cluster is unique but contains items that are like each other. Using the K-means algorithm, each image is grouped into a cluster whose mean is closest to it, resulting in a collection of clusters with related images. [68],



[69].

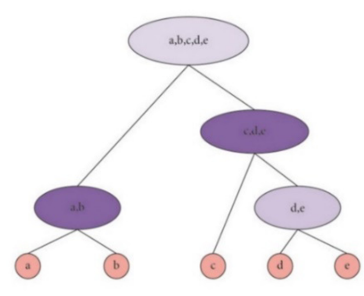
Fig. 5 Illustration of K-Means Clustering[70]

1) This study uses an algorithm and model to analyze the diseases of a corn leaf, it utilizes deep learning network and k-means clustering model. It uses 32 mean samples having an accurate result of the three common diseases found in the leaf with the method stated above. These studies techniques indicating that the approach it performed may be used in precision farming to safeguard crops as it detected an average accuracy of 93% for the diseases found in corn crop [69].

2) The paper composed by Muhamad Amirul Mohd Yusof and Ain Nazari designs an automated system to identify types of maize plant diseases using image processing. The method involves five stages: pre-processing, image segmentation, feature extraction, and classification. Histogram equalization and median filtering algorithms are used in pre-processing. Segmentation is performed using thresholding, masking, and k-means algorithms. In the feature extraction process, 13 features are extracted from the image for classification in a Support Vector Machine classifier. The system can identify maize plant diseases with an average accuracy of 97.53% for healthy leaves and 97.27% for diseased leaves [71].

B. Hierarchical Clustering

Hierarchical Clustering is an unsupervised machine learning method used to group related data points into clusters. It uses a dendrogram, a tree-like structure, to represent the dataset's clusters hierarchically. Each group can be further dividedinto distinct clusters or combined with other related groups based on their similarity. [72],



[73], [74].

Fig. 6 Illustration of Hierarchical Clustering Diagram[75]

1) The research [76] evaluated the responses of 11 hybrid baby corn genotypes to five levels of NaCl-induced salinity. The technique classifies genotypes and gauges how they react to various salinity conditions using multivariate parameters and hierarchical cluster analysis. It found that increasing salinity above 6 dS m–1 negatively affected all genotypes. The genotype Chang Daeng 18 had the poorest germination response, while PAC 571 was the most tolerant to salt stress. Physiological and biochemical parameters, such as free proline, membrane electrolyte leakage, and membrane stability index, as well as Na+ accumulation, were representative of salt stress. Salinity reduced leaf greenness but didn't affect net photosynthetic rate, stomatal conductance, and transpiration rate. Genotypes with higher salinity tolerance showed Na+ exclusion and a higher K+/Na+ ratio in leaves. The study suggests considering the relative susceptibility of genotypes to salinity in plant breeding programs, and that the selected salt-tolerant genotypes have potential for cultivation in salt-affected soils.
2) The research [77] investigated anthocyanins, natural pigments, in various parts of the maize plant, not just the grain. The purple corn variety Apache Red Cob was crossed with genetic stock 320 N, resulting in intense anthocyanin production in parts of the plant that are not usually pigmented. Hierarchical clustering is used in the study to classify profiles with similar compositions and assess the diversity of anthocyanin production in maize tissues Anthocyanin extracts from different parts of the plant were assessed, and a new pigment in anthers was identified as anthocyanidin 3-6″-phenylacetylglucoside. The study found that maize produces abundant anthocyanins in non-grain parts and that these extracts have diverse applications due to their varied pigment profiles and hues.

C. *Principal Component Analysis (PCA)*

Principal Component Analysis (PCA) is a statistical method that simplifies high-dimensional data by creating a set of linearly uncorrelated variables, known as principal components, from potentially correlated variables. It's used in various fields, including data mining, finance, psychology, computer vision, facial identification, and image compression. PCA combines redundant features to extract meaningful information from datasets. In the context of sweet corn genotypes, PCA helps understand variability by reducing data dimensions, allowing researchers to identify variables with the most variance and find the best-performing sweet corn lines. The principal components, ranging from 1 to 6, quantify the degree of variation.[78], [79].

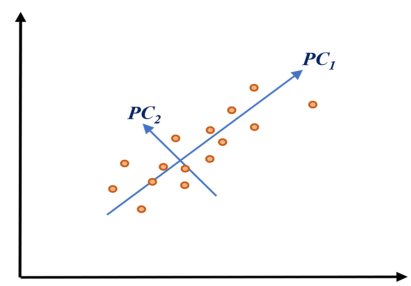


Fig. 7 Illustration of Hierarchical Clustering Diagram[80]

1) The research [81] evaluated the physical properties of maize seeds from a cold region in North China, using various agricultural material test benches. Parameters such as thousand-grain weight, moisture content, particle size, friction coefficients, angle of natural repose, coefficient of restitution, and stiffness coefficient were measured. Principal component and cluster analyses simplified the characteristic parameter index for judging the comprehensive score of maize seeds. Significant differences and correlations were found among the physical characteristics of the maize varieties. The first three principal component factors, representing over 80% of the information of the original eight parameters, were extracted. The physical characteristics of 15 kinds of maize seeds were evaluated and classified into four groups. The study provides a new approach for evaluating and analyzing the physical properties of agricultural materials and for screening and classifying food processing raw materials.

2) The study carried out by P. Divakara Sastry, E.V., Magudeeswari, and Th. Renuka Devi valuated plant nutrient traits in 12 baby corn genotypes using Principal Component Analysis (PCA) and cluster analysis. Variance analysis showed significant differences among genotypes for all traits except sugar content. PCA revealed the first three principal components accounted for 87.49% of variability, influenced by sugar and iron contents, and yield without husk. The genotypes were grouped into three clusters: Cluster-I (five genotypes) with higher iron content and yield without husk, Cluster-II (four genotypes) with higher potassium, phosphorous, and calcium contents, and Cluster-III (three genotypes) with higher sugar and phosphorous contents [82].
3) The research [83] examined corn gray leaf spot, corn rust, corn big spot, and healthy corn leaves. Image background segmentation was performed using the Otsu method, OpenCV morphological operation, and morphological transformation method to create an outline of the object and a mask. The difference set between the corn leaf and the background was used to get a complete corn leaf image. Principal Component Analysis (PCA) and Support Vector Machine (SVM) were applied to the processed image. With a penalty parameter C of 100 and a linear kernel in SVM, the classification accuracy for the four diseases was 90.05%, 92.64%, 91.23%, and 95.78% respectively.
4) The research [84] evaluated the adaptability and kernel quality of 10 maize hybrids. It examined grain yields, yield components, and grain quality characteristics. Kernel quality traits were mostly determined by the genotype, with significant differences among genotypes for all investigated traits. The BC hybrid stood out for traits like thousand kernel weight, ear width, number of kernels per ear, first ear height, and ear length. The BA hybrid had the highest values for plant height, day of silking, and day of tasseling. Positive correlations were found between traits like ear width, number of kernels per ear, and thousand kernel weight.
5) The study [85] conducted Principal Component Analysis on 26 sweet corn genotypes at the Tamil Nadu Agricultural University, Coimbatore in 2014. It observed eight quantitative and five qualitative traits to estimate variation and identify the best performing lines. Green cob yield was positively correlated with green cob length, girth, and number of kernel rows per cob, but negatively correlated with total sugar, sucrose, and starch. Principal component analysis showed variation by principal components 1 to 6. Clustering analysis grouped the 26 genotypes into eight clusters based on morphological traits rather than geographic origin. These diverse genotypes will be used for future breeding programs.

V. REVIEW OF MACHINE LEARNING MODELS FOR CORN QUALITY REGRESSION

A. Linear Regression

Regression is a supervised learning technique used for modeling continuous variables and making predictions. It can establish causal relationships between independent and dependent variables. Simple regression involves a one-to-one relationship between a single independent variable and the dependent variable. In contrast, multiple linear regression demonstrates a many-to-one relationship between several independent variables and one dependent variable, showing the influence of multiple factors on the outcome [86], [87].

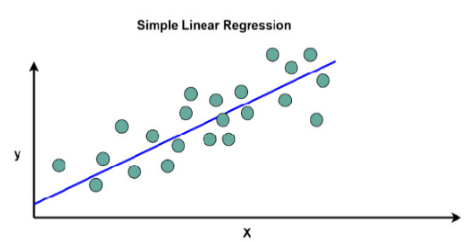


Fig. 8 Structure of Linear Regression[88]

1) [89] uses hyperspectral image technology and germination tests to predict the germination of sweet corn seeds. It analyzes 89 seeds and measures root and seedling length to assess seed vigor. Various regression methods are used to link hyperspectral seed features with germination results. The study finds that the method with the highest correlation coefficient gives the best prediction results for seedling and root length. The study concludes that hyperspectral technology can effectively predict seedling root length, suggesting a promising technique for predicting sweet corn seed germination.
2) The study [90] adapted an algorithm was adapted to predict starch, oil, and protein content of two maize cultivars using the APSIM-Maize model and a Three-Parameter Logistic model (3PLM). The APSIM-Maize model simulated the crop's phenology, growth, and grain protein, while the 3PLM computed the grain's starch and oil contents. The models were adjusted to fit experimental data and accurately predicted final starch and oil contents. The results suggest that combining APSIM with a 3PLM could be useful for predicting the protein, starch, and oil contents of maize grains.

B. *Decision Tree*

A decision tree is a supervised machine learning technique for classification and regression problems. It uses a tree structure where decisions are made at the leaves, and data is split at the nodes based on a parameter. Each path from root to leaf represents data separation steps leading to a result. This method provides a hierarchical representation of knowledge relationships [87], [91], [92], [93], [94], [95].

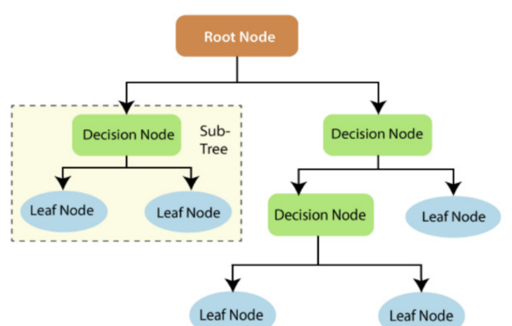


Fig. 9 Structure of Decision Tree[96]

1) The study used a decision tree model to classify diseases and pests in corn plants. However, due to ambiguity in discretizing predictor variables, the model's performance was unsatisfactory. To improve this, fuzzy membership functions were incorporated, significantly improving the model's performance. The proposed fuzzy model outperformed the decision tree model, effectively capturing data trends. [97].
2) IndojayaAgrinusa, a corn-based animal feed processing company, strives to maintain the quality of its products. Previous research used the Naive Bayes method in data mining techniques to determine corn quality standards, achieving a prediction accuracy rate of 82.3%. This study [98] used the Decision Tree (C 4.5) method, which increased the prediction accuracy to 86.2%.
3) The study [99] proposed a hybrid model for automated classification of three corn species in the Zea mays family, which is crucial for intelligent agriculture. The model used 12 different morphological features of corn and machine learning algorithms. Normal classification achieved a test score of 96.66% for Decision Tree, 97.32% for Random Forest, and 96.66% for Naive Bayes. However, the hybrid model achieved a 100% test score in all three algorithms, demonstrating its effectiveness in corn classification.
4) [100] proposes using the Random Forest, Neural Network, and Naive Bayes methods to classify diseases in corn plants, a staple food in Indonesia. Diseases can reduce corn production, and manual identification is inefficient. The dataset used is a collection of corn leaf images from the Madura Region, with four target classes: healthy, gray leaf spot, blight, and common rust. The Neural Network method provided better accuracy in classifying the corn leaf datasets, with an AUC value of 90.09%, classification accuracy of 74.44%, f1 score of 72.01%, precision of 74.14%, and recall of 74.43%.

C. *Ensemble Methods for Regression*

Ensemble learning is a supervised machine learning approach that uses several base-learners (algorithms) to generate decisions. These base-learners construct models, like classifiers or regressors, that generalize labeled examples. Predictions for fresh unlabeled examples can then be made using the generated model. In domains like healthcare, where even small gains in accuracy can have a big impact, ensemble approaches are especially useful because they usually yield more accurate results than a single model. There is no denying these techniques' efficacy, and when used appropriately, they can have enormous advantages. In fields such as healthcare, even the slightest improvement in the accuracy of machine learning algorithms can be extremely valuable [101], [102].

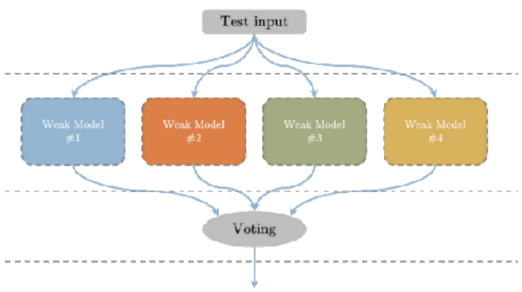


Fig. 10 Structure of Ensemble Method Model[103]

1) The study [104]uses a gradient boosting machine (GBM) learning model to predict aflatoxin contamination in corn. Utilizing historical corn contamination, meteorological, satellite, and soil property data, the model was created especially for Iowa. It was evaluated for two aflatoxin risk thresholds: 20-ppb and 5-ppb, achieving overall accuracies of 96.77% and 90.32% respectively. Corn contamination predictions were impacted by variables like soil-saturated hydraulic conductivity, aflatoxin risk indices in May, and the August satellite-acquired vegetative index. In order to guarantee the safety of food and feed, the study offers a potential technique for predicting aflatoxins in maize.
2) [105] The study used spectral data and machine learning methods to estimate leaf nitrogen in corn. It compared eight ML algorithms and found that gradient boosting and random forest were the best, with an 80% coefficient of determination. Adding vegetation indices to the spectral bands improved the results. The combination of SCCCI, NDRE, and red edge had the highest coefficient of determination in predicting leaf nitrogen content in corn.
3) The study [106] proposed a nondestructive detection of copper content in corn leaves using visible-near infrared spectroscopy. The data was preprocessed and the XGBoost predictive model was trained to predict copper content. The accuracy of the XGBoost and PLSR models was higher with the continuous removal spectrum and derivative of ratio spectroscopy. The model's robustness was tested with field-planted corn spectral data, showing potential for predicting heavy metal content in crops.
4) In the study [107] various machine learning algorithms were used to predict corn hybrid yields in the 2020 Syngenta Crop Challenge. The XGBoost model was the most accurate, with a root mean square error of 0.0524. This model was used to estimate yield performance for untested hybrids, identifying those with high predicted yields for potential breeding to increase corn production.

VI. COMPARATIVE ANALYSIS OF MACHINE LEARNING MODELS

Selecting the optimal machine learning model is crucial in artificial intelligence and data-driven decision making. These models analyze data to find patterns and relationships. Comparative analysis is key in assessing different models' effectiveness, benefits, and drawbacks. A thorough comparison study of machine learning models helps navigate options in different fields, enabling informed decisions for various datasets and real-world applications. The study explores the complexities of comparative analysis and the strengths and weaknesses of machine learning models [108], [109], [110], [111].

Support Vector Machines (SVM), Random Forest, and Neural Networks are versatile machine learning models with unique strengths and limitations. Their performance in specific tasks is assessed using evaluation metrics such as accuracy, precision, recall, and F1 score [112], [113]. In a similar vein, K-Means Clustering, Hierarchical Clustering, and Principal Component Analysis (PCA) are fundamental unsupervised learning models. K-Means excels at categorizing data into distinct clusters, Hierarchical Clustering creates a hierarchical hierarchy of associations, and PCA successfully decreases dimensionality while maintaining important information. These models are skilled in labeling data, which is required for later supervised learning tasks. They basically find intrinsic patterns in unlabeled data and then give labels to individual data points, making it easier to use supervised learning models [114].

We will examine the performance metrics utilized in their assessment, elucidating their metrics, advantages, and limitations.

A. Support Vector Machine (SVM)

Support Vector Machines (SVM) are sophisticated machine learning technique with several advantages and disadvantages. SVMs are among the most powerful and resilient classification and regression algorithms used in a wide range of applications [44]. Accuracy assesses overall correctness, precision assesses positive prediction accuracy, recall assesses the model's capacity to identify important occurrences, and the F1 Score finds a compromise between precision and recall [112].

SVM's success in high-dimensional spaces is one of its significant characteristics, making it well-suited for complicated datasets. It is suitable for a variety of classification challenges, including the diagnosis and prognosis of brain illnesses such as Alzheimer's, schizophrenia, and depression due to its high accuracies that are generalizable even in cases with high dimensionality [45]. Moreover, the usage of kernel functions demonstrates several distinct advantages in the categorization of tiny samples and has a greater recognition impact [115]. Additionally, it exhibits remarkable performance in both linear and non-linear data, showcasing its versatility, and is very resistant to overfitting, particularly in high-dimensional data situations, rendering it valuable in an array of applications [116].

SVM, however, is not without its problems. Its memory requirements can be high, especially when handling large datasets, which in some situations could limit its scalability. Support Vector Machines (SVMs) are susceptible to data noise and may perform poorly with data containing outliers or unnecessary features. They can be difficult to interpret, especially when using non-linear kernels, making it hard to understand decision boundaries in complex models. Despite these challenges, SVMs are commonly used in machine learning, particularly for high-dimensional and diverse datasets [44], [116].

B. Random Forest

Random Forest is an ensemble learning technique that combines multiple decision trees for accurate predictions. It's used in both regression and classification tasks in machine learning. A unique aspect of Random Forest models is their ability to provide insights into feature relevance, indicating how each feature contributes to the prediction performance. This aids in feature selection and model interpretation [117], [118], [119].

Random Forest is a machine learning technique that is effective for various issues. Its ensemble-based design and numerous decision trees make it resilient to overfitting and improve generalization. It can capture non-linear correlations, handle complex datasets, and manage high-dimensional data. However, it can be time-consuming and computationally expensive, especially with many trees or complex datasets. The models might not be as interpretable if the ensemble has many trees. The decision-making process can be elusive compared to simpler models. Also, feature sparsity in highly sparse data may limit its ability to recognize meaningful patterns [117], [119], [118]

To sum it up, the approach used in this study is random forest which can handle large and intricate datasets, therefore it confirms that random forest is an efficient and adaptable machine learning technique. Understanding feature importance enhances its interpretability and enables a thorough assessment of its performance using a variety of metrics. Random Forest may be computationally expensive during training and incomprehensible for complex models, yet it remains a popular choice for many applications where accuracy and robustness are essential.

C. Neural Networks

One kind of machine learning model that draws inspiration from the composition and functions of the human brain is the neural network [59]. The popularity of these models has increased, particularly when handling complex jobs and enormous amounts of data [120]. The correctness, completeness, and balance of predictions are measured using standard metrics for evaluating neural network performance, which include accuracy, precision, recall, and F1 score. However, because of the distinct nature of neural networks, new metrics such as the loss function, which measures the difference between predicted and actual values, are introduced. Improving model accuracy requires minimizing this loss during training [121].

One of neural networks' distinguishing characteristics is their capacity to excel at complex tasks, particularly when presented with huge and high-dimensional datasets [120]. Neural networks are capable of learning hierarchical and abstract representations from input data, making them well-suited for a wide range of applications like as image and speech recognition [122]. The automatic feature learning element is very useful since it eliminates the need for human feature engineering by allowing the model to recognize important patterns on its own. They have shown effective in solving a range of real-world issues, including complicated system modelling and handwriting recognition [59].

However, neural networks have significant shortcomings. They require large volumes of labelled data for good training, which might be difficult in situations where getting labelled data is time-consuming or impracticable. Furthermore, training time is an important factor, and advances in hardware, parallel computing, and optimization approaches try to alleviate these difficulties [120]. Furthermore, neural networks are prone to overfitting, particularly when the supplied data is inadequate, or the model design is very sophisticated.

Finally, neural networks are a strong paradigm in machine learning that may achieve outstanding performance on complex problems. Their capacity to automatically learn characteristics from data adds to their versatility, allowing them to be used in a variety of applications. Their efficacy, however, is dependent on vast volumes of labelled data, and the computing needs, especially for deep systems, require careful consideration.

D. K-Means Clustering

K-Means clustering is a popular machine learning approach that is classified as unsupervised learning. Its major goal is to divide a dataset into discrete groups, or clusters, based on data point similarity [68], [114]. Iterative methods are used by the algorithm to minimize the sum of squared distances between each data point and the centroid of its allocated cluster. This optimization criteria, called inertia, measures how well-formed the clusters are.

Several metrics are routinely used to evaluate the performance of K-Means clustering. A basic metric that the method aims to decrease is inertia, which measures the tightness of clusters. Lower inertia suggests more well-defined clusters. The silhouette score, which ranges from -1 to 1, provides information on cluster separation, with higher values indicating well-separated clusters [123], [124]. Furthermore, the Calinski-Harabasz Index assesses both cohesiveness and separation, assisting in the evaluation of cluster quality [125].

K-Means has major advantages that contribute to its popularity,including straightforward mathematical concepts, quick convergence, and ease of use [126]. Its scalability is especially useful when dealing with large volumes of data. Furthermore, K-Means is useful when clusters are well-separated and generally spherical in shape, as it performs better in such cases.

There are limitations to the algorithm, though. With the rise in data quantities, the traditional K-means method has encountered greater difficulty in analyzing huge data sets while meeting practical demands [126]. It is delicate when choosing the first cluster center placement, which, depending on the starting point, may have different results [127].

E. Hierarchical Clustering

Hierarchical Clustering known in its hierarchical representation of the clusters in a dataset [114]. It does not require a labelled dataset however is presents a dendrogram, a tree-structured graph that is used in heat maps to visualize the results of a hierarchical clustering calculation, is primarily used to determine the best way to assign objects to clusters [128].

The flexibility of Hierarchical Clustering to conform to the underlying structure of the data without the need for predefined cluster numbers is one of its key features. This flexibility is particularly helpful when working with datasets of varying complexity or when the number of clusters is unknown in advance [128].

Hierarchical clustering is a flexible method with no prerequisites, suitable for handling real-world data, the goal is to build a hierarchy of clusters. The efficacy of the suggested validity indexes in determining the true number of clusters and handling different types of data sets, including unbalanced data sets, is demonstrated by the experimental comparisons with the state-of-the-art validity indexes on fake and real-world data sets [129], [130].

F. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a machine learning technique used for dimensionality reduction. It transforms high-dimensional data into a lower-dimensional space, preserving significant information and the variance of the original dataset. This makes PCA useful for enhancing computational performance and overcoming the curse of dimensionality. PCA is used in various fields like neurology [131], quantitative finance [132], [133], and facial recognition [134], and is particularly effective in medical data correlation. It helps visualize complex, multidimensional data and understand underlying structures and trends. As an unsupervised feature extraction method, PCA can be applied even without labelled data [135], [136], [137], [138]

Principal Component Analysis (PCA) is a machine learning technique known for feature extraction and dimensionality reduction. However, it has limitations. PCA assumes linear correlations between variables, which can limit its performance with non-linear or non-Gaussian patterns. The primary components can be hard to interpret, and PCA might overlook significant traits with lower variance. The assumption of orthogonality across primary components might not always hold in real-world datasets, causing information overlap and inaccurate interpretation. Despite these limitations, PCA is valued for its ability to visualize high-dimensional data and preserve vital information. Users should be cautious of PCA's linear correlation assumptions and potential ambiguities when interpreting the transformed features [135], [83].

Overall, the paper discusses how machine learning, a key component of digitalization solutions, has gained significant attention in the digital arena. To assist readers in choosing the best learning algorithm to fulfill certain needs, this paper discusses the benefits and drawbacks of different algorithms from the standpoint of applications. With this, it provides profound understanding of the differences of machine learning models and its algorithms.

VII. RESULTS AND DISCUSSION

This comprehensive literature review highlights the various roles and functions of different machine learning models in assessing corn quality. When combined with IoT technologies, these models enhance several facets of corn farming, including quality assurance, real-time monitoring, and yield prediction. However, their effectiveness is contingent on specific tasks, data characteristics, and problem requirements. The machine learning models were divided into two categories: Supervised and Unsupervised, which included models like SVM, Random Forest, Linear Regression, CNN, ANN, K-Means Clustering, PCA, and Decision Tree. Each model was thoroughly discussed and analyzed, with a focus on comparing the performance of the methods found in the literature. The models' performance was assessed based on their usage, data characteristics, and problem requirements.

The results strongly indicate that the effectiveness of these machine learning models largely depends on specific tasks, data characteristics, and problem requirements. The conclusions are based on the existing literature, which primarily focuses on different machine learning models. These findings are important because corn is a globally significant crop, serving as a food source for both humans and animals and used in the production of biofuels, plastics, and other products. For future research, it is suggested to conduct a review comparing different models on a specific task.

VIII. CONCLUSION

This overview of the literature provides in-depth information on several models, a promising technique that can yield precise estimates in a range of applications, namely machine learning. A comprehensive review of studies dedicated to applications of machine learning in agriculture is presented in this paper. It focuses on the classification, regression, and clustering of corn crop evaluation. This paper also covers different approaches, techniques, and models of machine learning algorithm, including supervised and unsupervised learning. It also discusses and compare the advantages and limitations, trends and gaps, and patterns of each model based on its performance on a labelled and unlabelled dataset. The aim of this paper is to provide valuable insights into identifying various machine learning algorithm and choosing the appropriate model for certain problem situations needed to resolve, particularly in the corn crop quality evaluation.

REFERENCES

[1] A. Thayer *et al.*, "Integrating Agriculture and Ecosystems to Find Suitable Adaptations to Climate Change," vol. 8, no. 1, 2020.

[2] M. Lampridi, C. Sorenses, and D. Bochtis, "Agricultural Sustainability: A Review of Concepts and Methods," vol. 11, no. 8, 2019.

[3] L. Benos, A. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis, "Machine Learning in Agriculture: A Comprehensive Updated Review," vol. 21, no. 11, 2021.

[4] S. Mishra, D. Mishra, and G. H. Santra, "Applications of Machine Learning Techniques in Agricultural Crop Production: A Review Paper," vol. 9, no. 38, 2016.

[5] N. Gupta, A. K. Shrivastava, and P. Rawat, "Important Parameters Influencing Total Quality Management: A Comparative Study," vol. 174, Jun. 2020, doi: 10.1007/978-981-15-2647-3_8.

[6] J. Hsu, V. Kalesnik, and E. Kose, "What Is Quality?," pp. 44–61, 2019, doi: https://doi.org/10.1080/0015198X.2019.1567194.

[7] M. Paulsen, M. Singh, and V. Singh, "Measurement and Maintenance of Corn Quality," 2019, doi: https://doi.org/10.1016/B978-0-12-811971-6.00007-3.

[8] H. Velesaca, R. Mira, P. Suarez, C. Larrea, and A. Sappa, "Deep Learning based Corn Kernel Classification," *2020*, pp. 66–67.

[9] V. Smil, "Traditional Agriculture," 2022, doi: 10.4324/9780429038785-3.

[10] J. J. F. Montanez, "Smart Defect Detection and Sortation through Image Processing for Corn," pp. 1233–1237, 2020, doi: 10.1109/TENCON50793.2020.9293889.

[11] X. Li, B. Dai, and H. Sun, "Corn Classification System based on Computer Vision," vol. 11, no. 4, p. 591, Apr. 2019, doi: https://doi.org/10.3390/sym11040591.

[12] A. Gill, T. Kaur, and K. Devi, "Application of Machine Learning Techniques in Modern Agriculture: A Review," Aug. 2022, doi: https://doi.org/10.1145/3549206.3549255.

[13] R. Popli, D. Singh, R. Kumar, G. Saini, and K. Garg, "Role of Contemporary Techniques in Agriculture Development: A Systematic Review," pp. 1677–1681, Apr. 2022, doi: https://doi.org/10.1109/icacite53722.2022.9823819.

[14] H. JudeImmaculate, "Applications of Machine Learning Algorithms in Agriculture," 2020.

[15] K. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine Learning in Agriculture: A Review," *14 August 2018*, vol. 18, no. 8, Art. no. 8, doi: https://doi.org/10.3390/s18082674.

[16] M. Sadiku, "Machine Learning in Agriculture," 2018.

[17] Monica R. Mundada, "Machine Learning in Agriculture Application: Algorithms and Techniques," 2020.

[18] N. Yadav, S. M. Alfayeed, and A. Wadhawan, “MACHINE LEARNING IN AGRICULTURE: TECHNIQUES AND APPLICATIONS,” 2020.

[19] Abhinav Sharma, Arpit Jain, Prateek Gupta, and Vinay Chowdary, “Machine Learning Applications for Precision Agriculture: A Comprehensive Review,” 2021.

[20] Tatsiana Isakova, “InData Labs.”

[21] Sara Oleiro Araujo, Ricardo Silva Peres, Jose Cochicho Ramalho, Fernando Lidon, and Jose Barata, “Machine Learning Applications in Agriculture: Current Trends, Challenges, and Future Perspectives,” 2023.

[22] Jie Sun, Zulong Lai, L. Di, Ziheng Sun, Jianbin Tao, and Yonglin Shen, “Multilevel Deep Learning Network for County-Level Corn Yield Estimation in the U.S. Corn Belt,” 2020.

[23] Mailson Freire de Oliveira, Brenda Valeska Ortiz, Guilherme Trimer Morata, Andrés-F Jiménez, Glauco de Souza Rolim, and Rouverson Pereira da Silva, “Training Machine Learning Algorithms Using Remote Sensing and Topographic Indices for Corn Yield Prediction,” 2022.

[24] A. Ali *et al.*, “Machine learning approach for the classification of corn seed using hybrid features,” 2020.

[25] A. Călin, H. Muresan, and A. Coroiu, “Feasibility of Using Machine Learning Algorithms for Yield Prediction of Corn and Sunflower Crops Based on Seeding Date,” 2023.

[26] E. M. de Oliveira, D. S. Leme, B. H. G. Barbosa, M. P. Rodarte, and R. G. F. A. Pereira, “A computer VISION system for coffee beans classification based on computational intelligence techniques,” 2015.

[27] G. I. Gadotti, Nicacia Moraes, G. G. Silva, and R. D. C. M Monteiro, “PREDICTION OF RANKING OF LOTS OF CORN SEEDS BY ARTIFICIAL INTELLIGENCE,” 2022.

[28] B. D. Satoto and R. T. Wahyuningrum, “Corn Seed Classification Using Deep Learning as an Effort to Increase Corn Productivity,” 2021, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9651846

[29] Mohsen Shahhosseini, Guiping Hu, and S. Archontoulis, “Forecasting Corn Yield With Machine Learning Ensembles,” 2020.

[30] Daljeet S. Dhaliwal and Martin M. Williams, “Sweet corn yield prediction using machine learning models and field-level data,” 2023.

[31] Ersin Elbasi *et al.*, “Crop Prediction Model Using Machine Learning Algorithms,” 2023.

[32] Mohamed Kais Msakni, Anders Risan, and Peter Schütz, “Using machine learning prediction models for quality control: a case study from the automotive industry,” 2023.

[33] Morteza Nagahi, “Pattern recognition using machine learning for corn and soybean yield prediction,” 2020.

[34] Barkha Bhardwaj and Shivam Tiwari, “Exploring the Potential of Machine Learning in Agriculture: A Review of its Applications and Results,” 2022.

[35] Syamasudha Veeragandham and H. Santhi, “A Review on the Role of Machine Learning in Agriculture,” 2020.

[36] Sarah Condran, Michael Bewong, M. Islam, L. Maphosa, and Lihong Zheng, “Machine Learning in Precision Agriculture: A Survey on Trends, Applications and Evaluations Over Two Decades,” 2022.

[37] I. Attri, L. K. Awasthi, and T. P. Sharma, “Machine learning in agriculture: a review of crop management applications,” *1 July 2023*, pp. 1–41, doi: https://doi.org/10.1007/s11042-023-16105-2.

[38] Muhammad Hammad Saleem, Johan Potgieter, and Khalid Mahmood Arif, “Automation in Agriculture by Machine and Deep Learning Techniques: A Review of Recent Developments,” 2021.

[39] Dekera Kenneth Kwaghtyo and Christopher Ifeanyi Eke, “Smart farming prediction models for precision agriculture: a comprehensive survey,” 2022.

[40] S. Garcı´a-Lara and S. Serna-Saldivar, “Corn History and Culture,” pp. 1–18, 2019, doi: https://doi.org/10.1016/B978-0-12-811971-6.00001-2.

[41] M. Mishra, S. Sharma, and R. Sharma, “Corn (Zea mays) as a Nutrient Source and Diet: A Review,” 2021, doi: https://doi.org/10.9734/jpri/2021/v33i52b33629.

[42] G. Swapna, G. Jadesha, and P. Mahadevu, “Sweet Corn – A Future Healthy Human Nutrition Food,” vol. 9, no. 7, 2020, doi: https://doi.org/10.20546/ijcmas.2020.907.452.

[43] J. Sanchez, S. Magaña, C. López, and Ó. G. Brambila, “Agrotechnical evaluation of manual implements for corn planting,” 2018, [Online]. Available: https://www.scielo.org.mx/pdf/remexca/v9nspe21/2007-0934-remexca-9-spe21-4370-en.pdf

[44] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, “A comprehensive survey on support vector machine classification: Applications, challenges and trends,” vol. 408, pp. 189–215, Sep. 2020, doi: https://doi.org/10.1016/j.neucom.2019.10.118.

[45] D. Pisner and D. Schnyer, “Support vector machine,” pp. 101–121, 2020, doi: https://doi.org/10.1016/B978-0-12-815739-8.00006-7.

[46] R. Gandhi, “Support Vector Machine — Introduction to Machine Learning Algorithms.” [Online]. Available: https://medium.com/towards-data-science/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47

[47] Ang Wu *et al.*, “Classification of corn kernels grades using image analysis and support vector machine”.

[48] Yuslena Sari, Mutia Maulida, Razak Maulana, Johan Wahyudi, and Ahmad Shalludin, “Detection of Corn Leaves Nutrient Deficiency Using Support Vector Machine (SVM),” 2021.

[49] Jinghua Wang, Lei Yan, Fan Wang, and Shanshan Qi, “SVM Classification Method of Waxy Corn Seeds with Different Vitality Levels Based on Hyperspectral Imaging,” 2021.

[50] Peng Xu, Ranbing Yang, Tiwei Zeng, Jian Zhang, Yunpeng Zhang, and Qian Tan, “Varietal classification of maize seeds using computer vision and machine learning techniques,” 2021.

[51] H. Tyralis, G. Papacharalampous, and A. Langousis, “A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources,” vol. 11, no. 5, p. 910, Apr. 2019, doi: https://doi.org/10.3390/w11050910.

[52] E. Y. Boateng, J. Otoo, and D. A. Abaye, “Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review,” vol. 8, pp. 341–357, Nov. 2020, doi: https://doi.org/10.4236/jdaip.2020.84020.

[53] Ms. Deepika Chauhan, Dr. Ranjan Walia, Dr. Chaitanya Singh, Dr. M.Deivakan, and Mr. Makhan Kumbhkar, “Detection of Maize Disease Using Random Forest Classification Algorithm,” 2021.

[54] Shahid Nawaz Khan, Dapeng Li, and Maitiniyazi Maimaitijiang, “A Geographically Weighted Random Forest Approach to Predict Corn Yield in the US Corn Belt,” 2022.

[55] Junfeng Gao, David Nuyttens, Peter Lootens, Yong He, and Jan G. Pieters, “Recognising weeds in a maize crop using a random forest machine-learning algorithm and near-infrared snapshot mosaic hyperspectral imagery,” 2018.

[56] Hwang Lee, Jinfei Wang, and Brigitte Leblon, “Using Linear Regression, Random Forests, and Support Vector Machine with Unmanned Aerial Vehicle Multispectral Images to Predict Canopy Nitrogen Weight in Corn,” 2020.

[57] Aeri Rachmad *et al.*, “Classification of Diseases on Corn Stalks using a Random Forest based on a Combination of the Feature Extraction (Local Binary Pattern and Color Histogram),” 2023.

[58] Dongdong Ma, Liangju Wang, Libo Zhang, Zhihang Song, Tanzeel U. Rehman, and Jian Jin, “Stress Distribution Analysis on Hyperspectral Corn Leaf Images for Improved Phenotyping Quality,” 2020.

[59] Chirag, “Overview of Neural Network,” Jun. 2022, doi: https://doi.org/10.48175/ijarsct-4851.

[60] N. McCullum, “Deep Learning Neural Networks Explained in Plain English,” DEEP LEARNING. [Online]. Available: https://www.freecodecamp.org/news/deep-learning-neural-networks-explained-in-plain-english/

[61] N. Kriegeskorte and T. Golan, “Neural network models and deep learning,” vol. 29, no. 7, 2019.

[62] Jun Zhang, Limin Dai, and Fang Cheng, “Corn seed variety classification based on hyperspectral reflectance imaging and deep convolutional neural network,” 2020.

[63] Abdul Waheed, Muskan Goyal, Deepak Gupta, Ashish Khanna, Abdul Ella Hassanien, and Hari Mohan Pandey, "An optimized dense convolutional neural network model for disease recognition and classification in corn leaf," 2020.
[64] Linbai Wang, Jingyan Liu, Jun Zhang, Jing Wang, and Xiafei Fan, "Corn Seed Defect Detection Based on Watershed Algorithm and Two-Pathway Convolutional Neural Networks," 2022.
[65] B. D. Satoto, R. T. Wahyuningrum, and B. K. Khotimah, "Classification of Corn Seed Quality Using Convolutional Neural Network with Region Proposal and Data Augmentation," 2023.
[66] Prakruti Bhatt, Sanat Sarangi, Anshul Shivhare, Dineshkumar Singh, and Srinivasu Pappula, "Identification of Diseases in Corn Leaves using Convolutional Neural Networks and Boosting," 2019.
[67] Neven Voca, Lato Pezo, Anamarija Peter, Danijela Suput, Bilijana Loncar, and Tajana Kricka, "Modelling of corn kernel pre-treatment, drying and processing for ethanol production using artificial neural networks," 2021.
[68] A. A. Aldino, D. Darwis, A. T. Prastowo, and C. Sujana, "Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency," 2021, doi: 10.1088/1742-6596/1751/1/012038.
[69] H. Yu *et al.*, "Corn Leaf Diseases Diagnosis Based on K-Means Clustering and Deep Learning," *15 October 2021*, vol. 9, pp. 143824–143835, doi: 10.1109/ACCESS.2021.3120379.
[70] Murad Abusubaih, "Intelligent Wireless Networks: Challenges and Future Research Topics," 2022.
[71] M. A. M. Yusof and A. Nazari, "The Disease Detection forMaize-Plant using K-Means Clustering," vol. 2, no. 2, pp. 834–841, Oct. 2021, doi: https://doi.org/10.30880/eeee.2021.02.02.100.
[72] Aditya, "What is Hierarchical Clustering?," Hierarchical Clustering: Applications, Advantages, and Disadvantages.
[73] Sonoo Jaiswa, "Java Point," Hierarchical Clustering in Machine Learning.
[74] Eric J., "5 Use Cases and Practical Examples of Hierarchical Clustering."
[75] Wei Guo, Lihua Liu, Yao Yao, and Tong Shen, "Refinement Method of Evaluation and Ranking of Innovation and Entrepreneurship Ability of Colleges and Universities Based on Optimal Weight Model," 2022.
[76] A. T. M. Tanjimul Islam *et al.*, "Salt tolerance of hybrid baby corn genotypes in relation to growth, yield, physiological, and biochemical characters," vol. 147, pp. 808–819, Jul. 2022, doi: https://doi.org/10.1016/j.sajb.2022.03.023.
[77] M. Paulsmeyer, K. Vermillion, and J. Juvik, "Assessing the diversity of Anthocyanin Composition in Various Tissues of Purple Corn," vol. 201, Sep. 2022, doi: https://doi.org/10.1016/j.phytochem.2022.113263.
[78] Priya Pedamkar, "What is Principal Component Analysis," Principal Component Analysis.
[79] A. Thanga Hemavathy, "PRINCIPAL COMPONENT ANALYSIS IN SWEET CORN (ZEA MAYS L. SACCHARATA.)," 2020.
[80] Mohsen Hesami and A. Maxwell P. Jones, "Application of artificial intelligence models and optimization algorithms in plant cell and tissue culture," 2020.
[81] H. Tang, C. Xu, Y. Jiang, J. Wang, Z. Wang, and L. Tian, "Evaluation of Physical Characteristics of Typical Maize Seeds in a Cold Area of North China Based on Principal Component Analysis," vol. 9, no. 7, p. 1167, Jul. 2021, doi: https://doi.org/10.3390/pr9071167.
[82] P. Magudeeswari, E. V. Divakara Sastry, and Th. Renuka Devi, "Principal component (PCA) and cluster analyses for plant nutrient traits in baby corn (Zea mays L.)," vol. 53, no. 3, pp. 353–357, 2019, doi: 10.18805/IJARe.A-5042.
[83] Z. Liu, Z. Du, Y. Peng, T. Ming, X. Liu, and W. Chen, "Study on Corn Disease Identification Based on PCA and SVM," pp. 661–664, 2020, doi: 10.1109/ITNEC48623.2020.9084689.
[84] G. E. Orhun, "Investigation of Agronomic and Kernel Quality Traits of Registered Maize Varieties using Principal Component Biplot Analysis," vol. 65, no. 2, p. 7, Oct. 2020.
[85] A. T. Hmeavathy, "PRINCIPAL COMPONENT ANALYSIS IN SWEET CORN (ZEA MAYS L. SACCHARATA.)," vol. 45, no. 4, pp. 264–268, Mar. 2020.
[86] D. H. Maulud and A. M. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," vol. 1, no. 4, pp. 140–147, 2020.
[87] S. Ray, "A Quick Review of Machine Learning Algorithms," pp. 35–39, 2019, doi: doi:10.1109/COMITCon.2019.8862451.
[88] R. Uddin, "COVID-19 Pandemic Data Analysis and Prediction Using Machine-Learning Algorithms," Sep. 2023, doi: 10.31224/3226.
[89] Huawei Cui, Zhishang Cheng, Peng Li, and Aimin Miao, "Prediction of Sweet Corn Seed Germination Based on Hyperspectral Image Technology and Multivariate Data Regression," 2020.
[90] Maryam Rahimi Jahangirlou, Julien Morel, Gholam Abbas Akbari, Iraj Alahdadi, Saeid Soufizadeh, and David Parsons, "Combined use of APSIM and logistic regression models to predict the quality characteristics of maize grain," 2023.
[91] F.-J. Yang, "An Extended Idea about Decision Trees," pp. 349–354, Dec. 2019, doi: 10.1109/CSCI49370.2019.00068.
[92] J. Liang, Z. Qin, S. Xiao, L. Ou, and X. Lin, "Efficient and secure decision tree classification for cloud-assisted online diagnosis services," pp. 1632–1644, Jun. 2019, doi: 10.1109/TDSC.2019.2922958.
[93] A. M. Abdulazeez and A. S. Eesa, "Intrusion Detection and Attack Classifier Based on Three Techniques: A Comparative Study," Jan. 2021.
[94] A. Suresh, R. Udendhran, and M. Balamurgan, "Hybridized neural network and decision tree based classifier for prognostic decision making in breast cancers," vol. 24, pp. 7947–7953, May 2019.
[95] D. Kumar and Priyanka, "Decision tree classifier: a detailed survey," vol. 12, no. 3, pp. 246–269, 2020.
[96] B. T. Jijo and A. M. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," vol. 2, no. 1, pp. 20–28, 2021.
[97] Yulia Resti, Chandra Irsan, Muflika Amini, Irsyadi Yani, Rossi Passarella, and Des Alwine Zayantii, "Performance Improvement of Decision Tree Model using Fuzzy Membership Function for Classification of Corn Plant Diseases and Pests," 2022.
[98] A. Saleh and M. Maryam, "Determination of Corn Quality using the Decision Tree of C 4.5 Algorithm," 2019.
[99] Emre Avuçlu, Şakir Taşdemir, and Murat Köklü, "A new hybrid model for classification of corn using morphological properties," 2023.
[100] A Ubaidillah, Eka Malasari Rochman, Doni Abdul Fatah, and Aeri Rachmad, "Classification of Corn Diseases using Random Forest, Neural Network, and Naive Bayes Methods," 2022.
[101] O. Sagi and L. Rokach, "Ensemble learning: A survey," vol. 8, no. 4, 2018, doi: 10.1002/widm.1249.
[102] V. Bolón-Canedo and A. Alonso-Betanzos, "Ensembles for feature selection: A review and future trends," Dec. 2019, doi: https://doi.org/10.1016/J.INFFUS.2018.11.008.
[103] A. Anwar, "What are Ensemble methods in Machine Learning?," Towards Data Science. [Online]. Available: https://towardsdatascience.com/what-are-ensemble-methods-in-machine-learning-cac1d17ed349
[104] Emily H. Branstad-Spates *et al.*, "Gradient boosting machine learning model to predict aflatoxins in Iowa corn," 2023.
[105] Razieh Barzin, Hamid Kamangir, and Ganesh C. Bora, "Comparison of Machine Learning Methods for Leaf Nitrogen Estimation in Corn Using Multispectral UAV Images," 2021.
[106] Bing Wu, Keming Yang, Yanru Li, and Jiale He, "Hyperspectral Inversion of Heavy Metal Copper Content in Corn Leaves Based on DRS–XGBoost," 2023.
[107] Farnaz Babaie Sarijaloo, Michele Porta, Bijan Taslimi, and Panos M. Pardalos, "Yield performance estimation of corn hybrids using machine learning algorithms," 2021.
[108] A. Sharma, A. Jain, P. Gupta, and Vi. Chowdary, "Machine Learning Applications for Precision Agriculture: A Comprehensive Review," vol. 9, pp. 4843–4873, Dec. 2020, doi: 10.1109/ACCESS.2020.3048415.
[109] S. Durga and K. U. Rani, "A Perspective Overview on Machine Learning Algorithms," Dec. 2019, doi: https://doi.org/10.1007/978-3-030-46939-9_30.

[110] P. Ozoh, A. Adigun, and L. Omotosho, "A Comparative Analysis of Machine Learning Techniques," 2020.
[111] S. Sherin, M. U. Khan, and M. Z. Iqbal, "A Systematic Mapping Study on Testing of Machine Learning Programs," Jul. 2019, doi: https://doi.org/10.48550/arXiv.1907.09427.
[112] R. Yacouby, and D. Axman, "Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models," Nov. 2020, doi: 10.18653/v1/2020.eval4nlp-1.9.
[113] J. C. Obi, "A comparative study of several classification metrics and their performances on data," Feb. 2023, doi: https://doi.org/10.30574/wjaets.2023.8.1.0054.
[114] M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, and A. Aljaaf, "A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science," 2020, doi: 10.1007/978-3-030-22475-2.
[115] Ang Wu *et al.*, "Classification of corn kernels grades using image analysis and support vector machine," 2018.
[116] S. Ghosh, A. Dasgupta, and A. Swetapadma, "A Study on Support Vector Machine based Linear and Non-Linear Pattern Classification," pp. 24–28, 2019, doi: 10.1109/ISS1.2019.8908018.
[117] M. I. Prasetiyowati, N. Maulidevi, and K. Surendro, "Feature selection to increase the random forest method performance on high dimensional data," vol. 6, pp. 303–312, Nov. 2020, doi: https://doi.org/10.26555/ijain.v6i3.471.
[118] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," vol. 134, pp. 93–101, Nov. 2019, doi: https://doi.org/10.1016/j.eswa.2019.05.028.
[119] P. Probst, M. N. Wright, and A.-L. Boulesteix, "Hyperparameters and tuning strategies for random forest," Jan. 2019, doi: https://doi.org/10.1002/widm.1301.
[120] M. Islam, G. Chen, and S. Jin, "An Overview of Neural Network," vol. 5, no. 1, pp. 7–11, Shangzhu, doi: 10.11648/j.ajnna.20190501.12.
[121] S. Jandial, A. Chopra, M. Sarkar, P. Gupta, and B. Krishnamur, "Retrospective Loss: Looking Back to Improve Training of Deep Neural Networks," Jun. 2020, doi: https://doi.org/10.1145/3394486.3403165.
[122] S. Sakib, N. Ahmed, A. J. Kabir, and H. Ahmed, "An Overview of Convolutional Neural Network: Its Architecture and Application," 2019, doi: 10.20944/preprints201811.0546.v4.
[123] A. R. Mamat, F. S. Mohamed, M. A. Mohamed, N. M. Rawi, and M. I. Awang, "Silhouette index for determining optimal k-means clustering on images in different color models," Apr. 2018, doi: https://doi.org/10.14419/ijet.v7i2.14.11464.
[124] K. R. Shahapure and C. K. Nicholas, "Cluster Quality Analysis Using Silhouette Score," pp. 747–748, Oct. 2020, doi: https://doi.org/10.1109/DSAA49011.2020.00096.
[125] X. Wang and Y. Xu, "An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index," vol. 569, no. 5, 2019, doi: 10.1088/1757-899X/569/5/052024.
[126] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," vol. 2, no. 2, pp. 226–235, Jun. 2019, doi: 10.3390/j2020016.
[127] B. Chong, "K-means clustering algorithm: a brief review," vol. 4, no. 5, pp. 37–40, Sep. 2021, doi: 10.25236/AJCIS.2021.040506.
[128] P. Govendor and V. Sivakumar, "Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019)," vol. 11, no. 1, pp. 40–56, Jan. 2020, doi: https://doi.org/10.1016/j.apr.2019.09.009.
[129] Q. Xu, Q. Zhang, J. Liu, and B. Luo, "Efficient synthetical clustering validity indexes for hierarchical clustering," vol. 151, Aug. 2020, doi: https://doi.org/10.1016/j.eswa.2020.113367.
[130] S. Badillo *et al.*, "An Introduction to Machine Learning," vol. 107, no. 4, pp. 871–885, Mar. 2020, doi: https://doi.org/10.1002/cpt.1796.
[131] X. Zhuang, Z. Yang, and D. Cordes, "A technical review of canonical correlation analysis for neuroscience applications," vol. 41, no. 13, Sep. 2020, doi: 10.1002/hbm.25090.
[132] M. Carreras-Simó and G. Coenders, "Principal component analysis of financial statements. A compositional approach," vol. 29, pp. 18–37, Jun. 2020, doi: http://dx.doi.org/10.46661/revmetodoscuanteconempresa.3580.
[133] P.-F. Dai, X. Xiong, and W.-X. Zhou, "A global economic policy uncertainty index from principal component analysis," vol. 40, May 2021, doi: https://doi.org/10.1016/j.frl.2020.101686.
[134] M. Wang, L. Song, K. Sun, and Z. Jia, "F-2D-QPCA: A Quaternion Principal Component Analysis Method for Color Face Recognition," vol. 8, pp. 217437–217446, 2020, doi: 10.1109/ACCESS.2020.3041847.
[135] G. Rebala, A. Ravi, and S. Churiwala, "Principal Component Analysis," 2019, doi: https://doi.org/10.1007/978-3-030-15729-6_12.
[136] H. Abdi and L. J. Williams, "Principal Component Analysis," 2019.
[137] D. N. H. Thanh, N. N. Hien, P. V. B. Surya, U. Erkan, and A. Khamparia, "Adaptive thresholding skin lesion segmentation with gabor filters and principal component analysis," vol. 1125, 2020, doi: https://doi.org/10.1007/978-981-15-2780-7_87.
[138] J. Zhang, W. Cui, X. Guo, and B. Wang, "Classification of digital pathological images of non-Hodgkin's lymphoma subtypes based on the fusion of transfer learning and principal component analysis," vol. 47, no. 9, Jun. 2020, doi: 10.1002/mp.14357.

