IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 13 May 2024, accepted 21 May 2024, date of publication 23 May 2024, date of current version 4 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3404651

RESEARCH ARTICLE

Internet of Things (IoT)-Enabled Machine Learning Models for Efficient Monitoring of Smart Agriculture

MOHAMMAD ALDOSSARY[1], HATEM A. ALHARBI[2], AND CH ANWAR UL HASSAN[3]

[1]Department of Computer Engineering and Information, College of Engineering, Prince Sattam Bin Abdulaziz University, Wadi Al-Dawasir 11991, Saudi Arabia
[2]Department of Computer Science, College of Computer Science and Engineering, Taibah University, Madinah 42353, Saudi Arabia
[3]Department of Creative Technologies, Air University, Islamabad 44000, Pakistan

Corresponding author: Mohammad Aldossary (mm.aldossary@psau.edu.sa)

This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2024/R/1445).

ABSTRACT Advancements in the agricultural sector are essential because the need for food is rising as the population of the world is expanding day by day. Traditional agricultural practices are not able to fulfill these needs. Furthermore, these practices are manual and are not optimized resulting in the wastage of resources, which is not suitable for resource-constrained agricultural environments. Besides this, the Internet of Things (IoT) network is playing an important role in the modern farming system. In this paper, we introduce an innovative IoT-enabled hybrid model for smart agriculture, integrating Machine Learning (ML) and Artificial Intelligence (AI) algorithms to provide a cost-effective and reliable decision-making system. Furthermore, we introduce a robust anomaly detection mechanism while applying the capabilities of Multilayer Perceptron (MLP), Naïve Bayes, and Support Vector Machine (SVM) on the dry beans' dataset. Hybrid models, combining neural networks with Random Forest and SVM, were also explored for anomaly detection in the dataset. Furthermore, deep learning models known as MobileNetV2, VGG16, and InceptionV3 are used for the classification of soil type datasets. The hybrid deep learning models were also developed, incorporating InceptionV3 with Long Short-Term Memory (LSTM) and VGG16 with fully connected dense layers. Two types of data sets are used in this study, which are the dry beans dataset (2021) and soil type Dataset (2024). Both datasets contain images. The ML techniques are applied to these datasets for anomaly detection. The simulations results show that the classification performance of the MobileNetV2 model, it has an accuracy and recall of 0.97. It shows that the model can correctly identify the soil type around 97%. On the other hand, the hybrid model combining random forest and neural network achieved an accuracy of 92%, further validating the effectiveness of our approach. Furthermore, the SVM model achieves an impressive overall accuracy of 0.93. Additionally, this accuracy is further enhanced with the integration of SVM and neural networks. Similarly, the hybrid model combining inception V3 with the LSTM layer exhibits a notable accuracy of 0.91, highlighting its efficiency in accurately classifying various instances. Lastly, the hybrid model employing random forest and neural network architecture achieves a commendable accuracy of 92%.

INDEX TERMS Bean classification, deep learning, Internet of Things (IoTs), machine learning (ML), soil classification.

I. INTRODUCTION

The advent of the Internet of Things (IoTs) has revolutionized almost every field of life. In an IoT network, each object is equipped with sensor nodes, which are deployed in the environment for monitoring [1]. Furthermore, these IoT devices are enabled to collect and share data on their own in remote and dispersed areas. The IoT devices have profound implications in almost every field of life, e.g., smart cities, vehicular network systems, wireless sensor networks,

The associate editor coordinating the review of this manuscript and approving it for publication was Maurice J. Khabbaz.

agriculture, healthcare, etc. IoT devices have the potential to transform ordinary objects into smart devices that not only operate on their own but also provide an efficient working symmetry [2]. Many studies discuss that IoT devices can be integrated with Artificial Intelligence (AI) for the creation of intelligent systems that can automate the process of decision-making. This Artificial Intelligence of IoT (AIoT) facilitates the network to collect and process a large amount of data without the involvement of any monitoring party [3]. The nodes in the AIoT network collect real-time data while utilizing the capabilities of IoTs and smart decision-making is ensured with the AI models, which ultimately reduces human intervention and leads to an efficient, responsive, and reliable system [4].

Many solutions are proposed by the authors for agriculture management and livestock management. The authors in [5] propose a mechanism to identify paddy leaf images and apply color and pattern analysis for the detection of deficiencies in them. Furthermore, the authors in [6] use IoT devices and cloud computing to aid agricultural forming. However, these traditional solutions are not efficient and require a huge amount of resources, high labor costs, and excessive power consumption [5], [6]. The applications of AIoT are very dispersed and vast. In recent times, researchers have made efforts to explore concepts of AIoT in smart agriculture to solve the aforementioned issues. The capabilities of AIoTs facilitate the agriculture stakeholders in sensing the data from fields and the automated decision-making process without any human involvement, which results in a cost-effective solution for agriculture and soil management [7]. The AIoT devices are used in agriculture to monitor and control agriculture parameters to enhance the productivity of soil and the efficacy of the overall process. Precision agriculture is a novel technique that uses modern information technologies for the optimization of crop production. The high-resolution data from multiple sources is used to utilize crop management operations for intelligent decision-making. The IoT devices are used to fill the gap in supply and demand in the agriculture sector, which ultimately ensures environmental sustainability with high crop yield and profitability. Some key areas need to be investigated in precision agriculture e.g., pest control, nutrient management, safe store management, and water management [8], [9], [10].

The integration of Machine Learning (ML) algorithms with IoT devices automates the functionalities of IoT devices, which are deployed in the agricultural field for monitoring water storage, nutrient management, and pest control. Furthermore, ML algorithms are used to analyze dispersed and vast volumes of datasets that are collected by IoT sensor nodes, which ultimately helps in providing a cost-effective and efficient solution for agricultural activities. The ML algorithms outperform all traditional methods and show the potential to revolutionize agricultural practices. Different countries in the world are facing severe challenges due to lack of water and agricultural experts. Also, countries like Saudi Arabia, Pakistan, India, and Uganda are not able to optimize their resources in the agricultural industry [13]. These countries rely on monsoon and much of their water is wasted during irrigation due to a limited portion of cultivated lands. It is the reason that there is a need to propose an efficient solution that not only enhances agriculture efficiency but also opens the doors for new technologies in these countries [14], [15].

Caring for plant health is widely acknowledged as a labor-intensive and costly task. The aforementioned discussion highlights the importance role of AI models and IoT networks in the agricultural sector. However, the existing AI models are not able to enhance the efficiency of agricultural yield while simultaneously achieving network sustainability [13]. Furthermore, the agricultural sector is still using traditional practices that are not optimized and consume a large amount of resources, which is not suitable for resource-constrained agricultural sector. Lastly, the traditional AI algorithms used in the agricultural sector suffer from high dimensionality and variability of visual agricultural data, which leads to the incorrect prediction of crop yields and identification of diseases [14], [15].

To address these challenges, we propose an IoT-enabled hybrid model that integrates ML and AI algorithms in the agricultural sector. This approach emphasizes the overarching objectives of maximizing agricultural yields, ensuring profitability, and promoting environmental sustainability. These objectives are given as follows:

1. Develop a comprehensive hybrid model integrating ML and AI to bolster global food security through the classification of the dry bean dataset, ensuring accuracy and scalability.
2. Utilize Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Naïve Bayes classifiers to streamline crop selection and enhance yield, leveraging their individual strengths to optimize agricultural outcomes.
3. Employ principal component analysis to extract crucial features and refine hybrid classifiers, mitigating overfitting and computational overhead while maximizing predictive accuracy and efficiency.
4. Investigate the efficacy of ensemble methods in augmenting classification accuracy, exploring synergistic interactions among ML models to uncover novel strategies for improving agricultural decision-making processes.

The main contributions of this paper are given as follows:

1. Propose the integration of AI and IoT to develop a sophisticated system capable of automating decisions pertaining to agriculture and soil management, thereby enhancing operational efficiency and sustainability.
2. Employ MLP, Naïve Bayes, and SVM ML models for anomaly detection within the dry beans' dataset, harnessing high-resolution data for autonomous and reliable decision-making. This would optimize various aspects of crop production, including nutrient management, water utilization, and pest control.

3. Explore the efficacy of different deep learning models such as MobileNetV2, VGG16, and InceptionV3, alongside hybrid architectures like InceptionV3 and LSTM Layer, VGG16, and Dense Layer Classifier for soil type classification. These classifiers effectively address the challenges associated with accurately classifying soil types, offering a robust solution to the resource-intensive and complex task at hand.

The rest of the paper is as organized follows. The related work is discussed in Section II. The datasets and pre-processing are presented in Section III. The proposed models are given in Section IV. The performance of the proposed system models is evaluated in Section V. The limitations of the proposed work are presented in Section VI. The paper ends with the conclusion in Section VII.

II. RELATED WORK

Different studies are conducted to propose efficient mechanisms to integrate IoTs and AI with the agricultural sectors to enhance growth efficiency to increase the output yield. The authors in [16] propose an Artificial Neural Network (ANN) enabled scheme for agricultural land. In this scheme, the IoT devices are deployed in the agricultural field to monitor temperature, pressure, humidity, and CO2 concentration in the environment. The IoT devices sense the data from the environment and then send it to the cloud server. The cloud server processes the data, removes all redundancies from the data, and converts it into meaningful information. The ML algorithms are deployed in the cloud server that classifies the data for managing conditions for a particular type of plant.

The authors in [17] state that the population in each region is growing day by day. So, there is a need to modernize the functionalities of agriculture to fulfill the food demand of that respective region. The environmental conditions cannot be completely mitigated but some strategies based upon the capabilities of ML and artificial intelligence can be adapted to improve the quality and quantity of crop fields. To solve these aforementioned issues, the authors propose IoT devices and ML algorithms enabled for the prediction of crop yield. The data of current weather and historical crop yield are combined for the prediction of seasonal crop yields. Different parameters of weather like rainfall, temperature, soil moisture, etc., are considered for prediction. These parameters are sensed by IoT devices that are deployed in the agricultural field for monitoring. After sensing the data, it is sent to the cloud server for processing. The data is aggregated, and all the redundancies are removed from that data. In this proposed model, the perspectives of policymakers and farmers' context are considered into the account to propose a more efficient strategy. The data on weather that is collected by IoT devices is fed into the ML algorithm for the prediction. A robust prediction model is used in the network to detect and sense information from the field with a working prototype of the proposed system. The authors consider different factors to evaluate the proposed model. These performance factors are a comparison of yield predicted, predicted data, accuracy, mean absolute percentage error, and seasonal yield. The results show that the proposed model outperforms all existing schemes with a mean absolute percentage error of 0.339. Similarly, Yao et al. discuss that the number of agricultural plants and planted areas is increasing day by day [18].

There are some approaches in which unmanned aircraft are deployed in the field that use smart sensors for monitoring crop growth and managing plant diseases. One of the most fundamental crop plants is the orchid due to its economic value. However, environmental factors and diseases are two factors that badly affect the growth of orchid plants. Therefore, there is a need to adopt a strategy to control the disease in orchid plants. This can be achieved by an immediate diagnosis of diseases to effectively find the prevention and treatment of that disease. Therefore, the authors propose an edge computing-based deep neural network for detecting diseases in orchid plants. The functionalities of edge computing are integrated with the IoT and deep neural networks for the detection of plant disease. Furthermore, deep learning neural network ensures subsequent dynamic learning. Three parameters are considered for dynamically adjusting the deployment of the network. These parameters are the deep neural network model, the computing capabilities of edge nodes, and the network's current state. The regional feature extraction is performed by the edge nodes. While on the other hand, all the global features are extracted and managed by cloud computing. Plant disease recognition with high precision and accuracy is done with the integration of deep learning networks, which effectively utilize all available resources without large network overhead. The first goal of the system is to mitigate the harmful effects of light. The lighting conditions are changed, and the direction of illumination is also changed. Secondly, there may be some noise in the image collected by IoT devices, particularly, in the case of poor sampling channels and bad conditions of lightning. Due to this both training and testing samples of images will be deviated from their original nature. The second goal of the system is to mitigate the effects of noise in the data. Thirdly, it is very complicated to collect a good quality image in conditions where orchid leaves can move freely. Therefore, the authors propose a strong and adaptable deep neural learning model. Fourthly, the nature and symptoms of the orchid diseases in different types of orchid plants are very diverse. Lastly, when three three-dimensional images are converted into two-dimensional images then some information can be lost. The most significant issue is that it is very complex and challenging to use edge computing to extract the data of three-dimensional while simultaneously addressing the issue of large data volume and the huge amount of time required for computation.

Some studies are also conducted to monitor and evaluate the amount of supplied water to agricultural land while utilizing the capabilities of ML algorithms and IoT devices [19], [20]. These studies use IoT devices and intelligent automated systems for the identification of ground characteristics e.g., moisture content, soil temperature, air pressure, etc. These characteristics are used to predict the relative humidity in the environment. Moreover, the water temperature can be intelligently regulated, and surrounding environmental conditions are adapted while utilizing the capabilities of ML algorithms and IoT devices. The authors state that there are a lot of countries in the world that are only relying on monsoons for agricultural products with a limited scale of agricultural fields. Around 85% water of in these countries is used in irrigation and a huge amount of data is lost in irrigation due to inefficient and unreliable techniques. Different IoT-based solutions are provided by authors in [21], [22], and [23], aiming towards assisting farmers in fulfilling the gap between demand and supply of agricultural yields while guaranteeing reasonable profit and environmental preservation. The authors in [21] state that the agricultural sector is evolving day by day and becoming more data-centric and precise. Therefore, advanced technologies like IoT networks, ML, and artificial intelligence-based solutions are being proposed for smart agriculture to enhance crops' yield and profitability while simultaneously reducing the amount of irrigation waste. However, the existing solutions are not efficient and provide a costly automated solution with a large computational overhead, which is not suitable for the resource-constrained agricultural sector. Therefore, Akanksha et al. propose a hybrid ML model that is integrated with an IoT network for the prediction of crop yield. There are three phases in the proposed model, which are pre-processing of data, feature extraction from data, and classification. In the first phase, the data collected from different sources is pre-processed to remove the unwanted value and noise from the data. After removing all the errors and outliers the data is sent to the feature extraction phase. Feature extraction is performed from the data to identify the most important and necessary features for crop production. The main benefit of feature extraction is that it enables the ML algorithm to train faster. Furthermore, feature extraction is also helpful in enhancing the accuracy of the model after the selection of the proper subset and reducing overfitting. After feature extraction, the classification of soil is done using an adaptive k nearest neighbor classifier. This classifier is an advanced version of the k nearest neighbor classifier and enhances the performance of the previous version. This machine-learning algorithm is used to achieve good accuracy while utilizing a small amount of time. There are three tiers of classification in the proposed model. In the first tier, the soil nutrient data collected from IoT devices is used to estimate the quality of soil using an adaptive k nearest neighbor classifier. Then the score of soil quality along with other parameters associated with crop yield like temperature and rainfall are used as an input to an extreme learning machine for the prediction of crop yield. The authors use PYTHON to implement the proposed algorithm and evaluate its performance. Soil data, accuracy, root mean square error, mean square error, and mean square logarithmic error are some parameters that are used to evaluate the performance of the proposed model. Similarly, different studies are conducted to showcase the efficiency of ML algorithms. These studies provide enough evidence that the support vector regression and multilayer perceptron-based techniques outperform all traditional studies in terms of accuracy, mean square error, precision, recall, and F1 score [24], [25], [26].

The linear and non-linear agricultural data can be easily predicted while utilizing the capabilities of these aforementioned techniques. Some techniques are also proposed to provide a secure IoT network for monitoring agricultural fields. The authors in [23] state that it is very important for network security that only authorized persons should have access to the data. Therefore, the authors propose a signature-based authentication mechanism for the agricultural sector in which all the IoT devices are authenticated based on their credentials. The authentication of nodes is performed by utilizing the capabilities of digital signatures and elliptic curve cryptography. Furthermore, the proposed mechanism considers passwords, mobile devices, and biometrics for avoiding denial-of-service attacks in the network. The proposed mechanism is also robust against various kinds of security attacks in the network and provides a secure and efficient environment for sensing data from agricultural fields.

III. DATASETS AND PRE-PROCESSING

This study uses two types of data sets, which are the dry beans dataset (2021) and soil type Dataset (2024). Both datasets contain images. The ML techniques are applied to these datasets for anomaly detection.

A. DRY BEAN DATASET (2021)

The dry beans dataset (2021) is publically available; all the images of this dataset are taken using a high-resolution camera [27]. This dataset consists of samples that represent different types of dry beans. These beans are different from each other based on their sizes, shapes, and textures. In this section, we consider seven types of dry beans, which are Sira, Seker, Horoz, Dermason, Cali, Bombay, and Barbunya. We use a computer vision system to differentiate between the varieties of dry beans to obtain uniform seed classification. The dataset includes 13,611 images of 7 different beans. The anomalies in this dataset can be unusual shapes and sizes of dry beans that deviate to a large extent from normal dry beans. First, the dry beans are obtained from a computer vision system then segmentation and feature extraction are performed on this data, and a total of 16 features; 12 dimensions and 4 shape forms, were obtained from the grains.

B. SOIL TYPE DATASET (2024)

The soil type dataset consists of samples of different soil types. There are different features on the basis on which the soil samples are distinguished, e.g., color, moisture content, and soil composition. This dataset contains images, which are also taken from high-resolution cameras. There are around 144 labeled photos that show the diversity of soils. The images were labeled as Alluvial Soil, Clayey Soil, Laterite Soil, Loamy Soil, Sandy Loam and Sandy Soil. There may be different anomalies in the dataset e.g., soil samples can have atypical properties due to pollution in or other environmental factors.

C. DATASET INSIGHTS OF DRY BEAN DATASET (2021)

The following bar chart shows the distribution of different classes of beans with the dry beans' dataset.

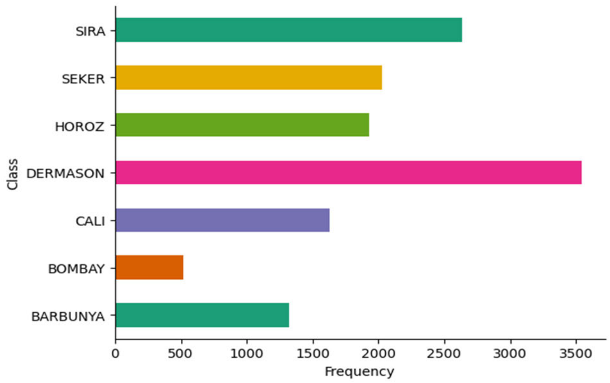

FIGURE 1. **Frequency of beans in the dataset.**

In Fig. 1, it is shown that the frequency of various classes is different in the dataset. DERMASON is the class, which has the highest frequency that exceeds 3000 instances. It can be suggested that DERMASON is the most commonly cultivated bean. The class that has the second highest frequency is SIRA, followed by SEKER. These both classes have frequencies of more than 2000 instances. Furthermore, the frequency of HOROZ and CALI is moderate, noticing the fact that the frequency of HOROZ is slightly higher than CALI. The two classes with the least frequencies are BOMBAY and BARBUNYA, where the frequency of BOMBAY is the least among all classes.

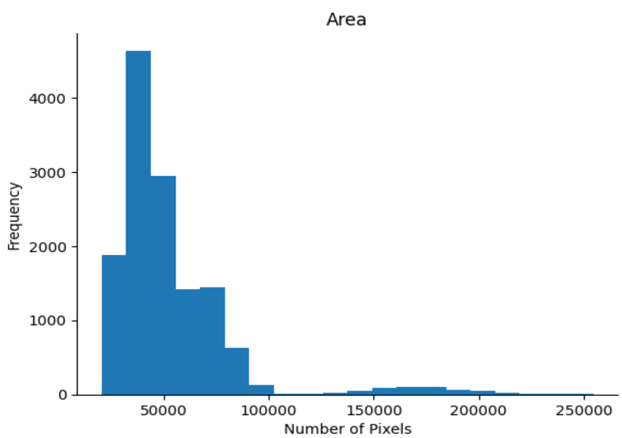

FIGURE 2. **Bean area distribution in the dry bean dataset.**

Here, one most important thing is that all classes are imbalanced, we utilized the capabilities of the weighted loss function for class balancing, which ultimately ensures that the model is not biased towards the most prevalent class. Fig. 2 shows the distribution of areas for each bean in pixel count. It can be observed that the area of most of the bean's ranges from 50,000 pixels and the bean with the highest frequency is falling in the smaller size range. This indicates that the beans that are most prevalent have smaller sizes or the method which is used for collecting the data of beans is tilted towards smaller-sized beans. There is an inverse relationship between the area size of beans and their frequency, which indicates the larger beans are least common in the dataset.

Fig. 3 shows the distribution of perimeter across each bean in pixel count. It can be observed that a large volume of beans lies in the lower perimeter range, particularly between 600 and 800 pixels.

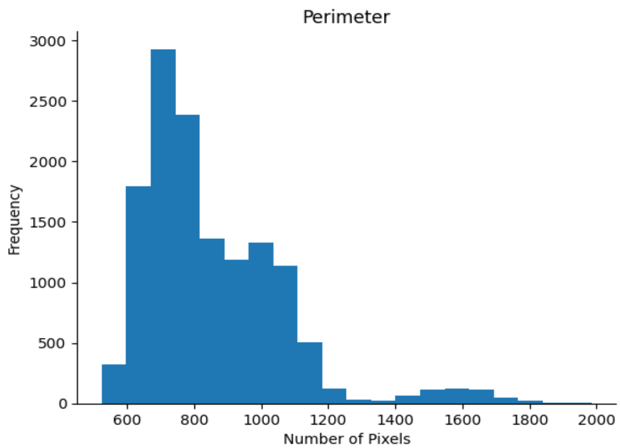

FIGURE 3. **Bean perimeter distribution in the dry beans' dataset.**

It means that there are a large number of beans that have smaller perimeters. When the perimeter of beans increases beyond 800 pixels, then the frequency of beans decreases. It indicates that beans with much larger or irregular shapes are less common. Fig. 4 showcases the heatmap of Pearson correlation between features of the dry beans' dataset. The heatmap represents the correlation between the features of beans having more strong positive correlations between perimeter and area and the strong negative correlation between area and shapefactor1.

Solidity and extent are two features that show a low correlation with other parameters, which ultimately indicates that their role is distinctive in classification. This pattern of correlation helps the feature selection for ML models, which ultimately helps to avoid redundancy in features and guides in the selection of kernels for the SVM classifier and feature engineering for the Naïve Bayes classifier.

D. DATASET INSIGHTS OF SOIL TYPE DATASET (2024)

The bar chart in Fig. 5 shows the distribution of sample counts among different classes of soil in the soil type dataset.

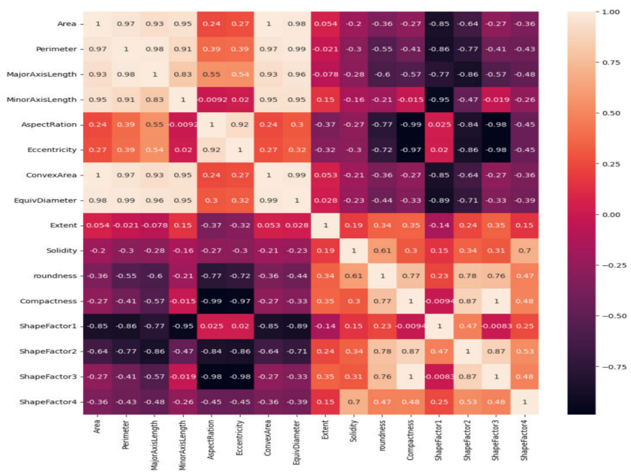

FIGURE 4. Pearson correlation coefficients between features of the dry beans' dataset.

It can be seen in the figure that the sandy soils have the highest frequency, which indicates that this soil is commonly sampled or prevalent soil in the dataset. Furthermore, the clayey and laterite soils have the second and third highest frequencies, respectively. On the other hand, loamy soil has a comparatively low frequency, which shows that this soil is less common or under-sampled in the given dataset. While alluvial soil has a moderate frequency, due to which, it can be indicated that alluvial soil is well sampled. The variation and disparity in the soil can lead to biases in the model training. Due to this, the model will perform well with the soil, which has more samples. To solve this issue, we utilize the weighted loss function, which ensures all types of soil are equally sampled and contributing equally to the learning process [28].

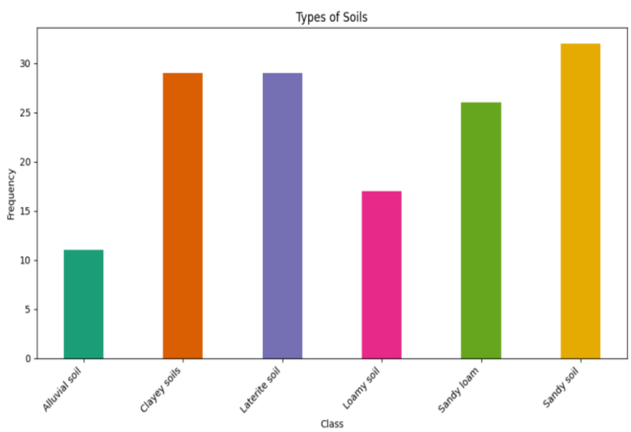

FIGURE 5. Frequency of sample counts across different soil classes.

Fig. 6 presents the visual comparison of pixel intensity of all six types of soils in the dataset. It can be observed from the fig. that alluvial soil is very famous for its fertility and is found in river basins. Furthermore, there is multimodal distribution that indicates complex composition with organic matter and moisture. It is also shown that the intensity value of clayey soil is low, and it reflects less light high retention of moisture, and finer particles.

The soil that has the highest concentration of iron and aluminum is laterite soil. This soil is recognizable due to its reddish color and has mid-range intensity. Moreover, it can also be observed that loamy soil has a balanced texture and is considered to be ideal soil for agriculture. Loamy soil has a lower frequency distribution, which shows it has fewer samples. On the other hand, sandy loam and sandy soil have larger sizes of particles and lower organic content and they correspond with dryer conditions and lighter colors.

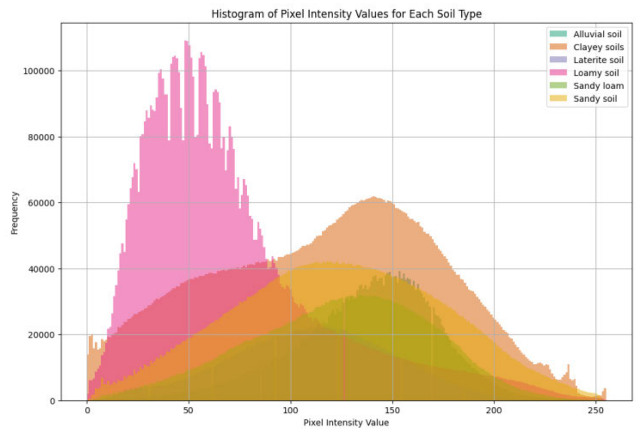

FIGURE 6. Pixel intensity values for various soil types.

E. DATA PREPROCESSING

In this section, we detect anomalies from the dry beans' dataset (2021) while utilizing the capabilities of MLP, Naïve Bayes, and SVM. For pre-processing of data, it is loaded in the environment, and it ensures alignment of all the features and labels. After that Z-score normalization is used to normalize the dataset, as given in (1) [29].

$$Z = (X - \mu)/\sigma \tag{1}$$

where μ represents the mean pixel intensity and is set to 0 and standard deviation is represented by σ and its value is 1 and X shows the value of the original feature value. Furthermore, the principal component analysis feature selection technique is used to identify the most informative features from anomaly detection. After that, the generalization of models is enhanced by utilizing data augmentation techniques known as random flipping and rotation of the bean images. In last, the dataset was split into testing and training datasets, the models MLP, Naïve Bayes, and SVM were trained by the training dataset, and the performance of the model was evaluated using the testing dataset.

On the other hand, deep learning models, namely MobileNetV2, VGG16, and InceptionV3 are used to detect anomalies from soil type Dataset (2024). In the data pre-processing step, the data of soil type is loaded, and it is ensured that all modalities are aligned correctly. Then uniformity of data is ensured by resizing the images into standard dimensions and applying skull stripping techniques that isolate the region of interest. After that, the Z-score normalization is applied to soil images to reduce contrast and variations in intensity. The images are flipped randomly on different planes e.g., horizontally and vertically to get the multi-view of the soil image data, the probability is set at 0.5. Furthermore, the noise is removed, and the quality of images is enhanced by utilizing the capabilities of the Gaussian blurring technique. In last, the dataset was split into testing and training datasets, the deep learning models MobileNetV2, VGG16, and InceptionV3 were trained by using the training dataset, and the performance of the model was evaluated using the testing dataset.

IV. PROPOSED IoT ENABLED HYBRID MODELS FOR SMART AGRICULTURE

In this section, we propose a mechanism in which IoT devices are deployed in the agricultural field for the monitoring of different parameters like soil moisture, humidity, temperature, nutrient level, etc. As shown in Fig. 7, IoT devices are equipped with different sensors and deployed in the agricultural field while ensuring comprehensive coverage of the field. It makes a significant advancement in smart agriculture and facilitates an accurate sensing of these parameters Furthermore, it also provides real-time data, which ultimately helps in intelligent decision-making very rapidly. The IoT devices sense the data from the environment and send it to the cloud servers.

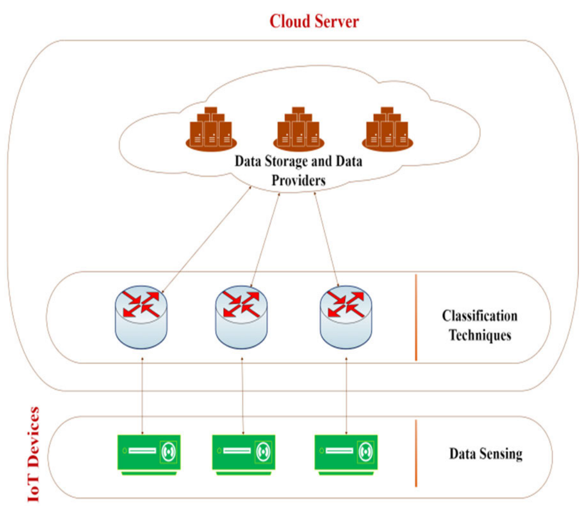

FIGURE 7. IoT enabled network for data sensing.

The cloud server acts as a centralized unit and contains very high storage and computational capabilities. The cloud server collects the data from IoT devices and aggregates this data. All the duplicative values and redundancies are removed from the data and stored in the cloud server for model training. The aggregated data on the cloud server is used for the training of deep and ML models. MLP, Naïve Bayes, and SVM ML models are trained and used for classifying the dry bean dataset while considering the shapes, textures, and sizes of these beans. While on the other hand, MobileNetV2, VGG16, and InceptionV3 are utilized for the classification of different soil types. The training of these models is done to identify patterns and properties in soil images and then they are used for classification to enhance crop production and land management. Fig. 8 represents the flow diagram that illustrates the sequential steps undertaken in this research, providing a visual overview of the research process.

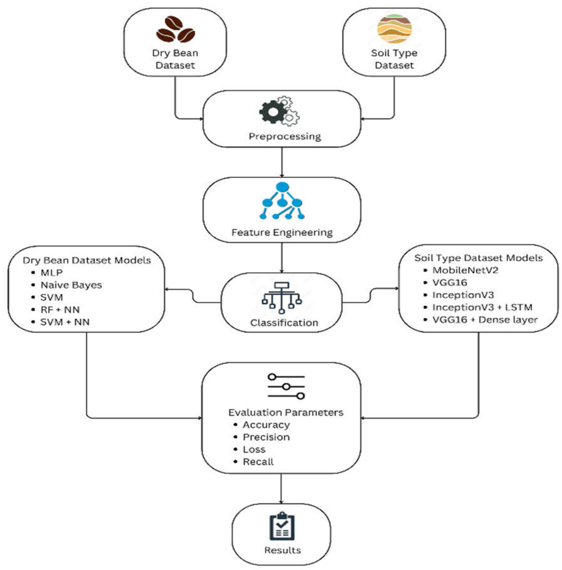

FIGURE 8. Methodology flow diagram.

A. PROPOSED MODEL FOR CLASSIFICATION OF DRY BEAN DATASET (2021)

The dry beans dataset was collected in 2021, it represents the dispersed properties of various bean types. As beans are one of the most used food crops worldwide; therefore, the study to increase their yield aligning with the objectives of global food security and sustainability is very important. There are different morphological features in the dataset, e.g., area, shape, and texture that play a significant role in computational analysis and predictive modeling [30]. The accurate classification of bean types is very important to ensure the standard sizes, textures, and shapes of beans that ultimately provide consistency in the products. Furthermore, this classification also helps in the selection of desired properties that ultimately enhance crop yield, and climate adaptability and provide efficient strategies for resistance to diseases.

In this section, we utilize SVM, MLP, and Naïve Bayes ML algorithms for the classification of the dry beans' dataset. SVM is an ML algorithm that provides efficient classification for the small size of the population. While MLP classifier is a type of neural network, which efficiently captures and models complex relationships in high-dimensional data. In last, Naïve Bayes uses the probabilistic approach and very efficient classifier for multi-class classification scenarios. All the above-mentioned properties of these three classifiers make them suitable for classifying the dry beans dataset. We use these three classifiers to achieve precision in the classification and compare the performances of these classifiers on the dry beans' dataset. It not only helps us in the identification of the most suitable classifier but also in exploring potential synergies in an ensemble approach.

B. FEATURE SELECTION AND EXTRACTION

It is very important to select the most prominent features and their extraction for data analysis. It helps in enhancing the performance and efficiency of predictive models. As the dataset of dry beans is multi-faceted and presents a multi-dimensional space it is very crucial to perform feature selection with this dataset. In this section, we isolate the most prominent and informative features from the dataset and reduce dataset redundancy. It is a very critical step to perform dimensionality reduction and it could result in overfitting and marginally increase the computational cost of deploying the model. We utilize Principal Component Analysis (PCA) and mutual information to remove redundancies from the dataset. We utilize the properties of the PCA method to transform original correlated features into linearly uncorrelated variables, which are called principal components. This method ensures that the first principal component has as high variability in the dataset as possible, which results in the reduced set of variables that contain the most significant information. These variables are responsible for explaining the variability in the data maximally [31]. The feature selection process marginally improves the performance of models. The risk of overfitting is reduced by reducing the number of input parameters, which ultimately helps the models to work efficiently without utilizing a large number of resources. Feature selection method PCA helps enhance the efficiency of the MLP model, which is a very resource-consuming classifier. On the other hand, we utilize the properties of feature selection by reducing the impact of independent assumption for the Naïve Bayes classifier [32]. For the dry beans' dataset, the above-mentioned PCA method of feature selection and extraction is helpful in the effective classification of beans by retaining the most prominent information in the features while simultaneously discarding irrelevant features that are not helpful in class separation. This method gives us the most optimized set of features that have intrinsic variances within bean types while simultaneously enhancing the operational efficiency of classifiers.

C. PROPOSED MODEL FOR CLASSIFICATION OF SOIL TYPE DATASET (2024)

In this section, we use deep learning models known as MobileNetV2, VGG16, and InceptionV3 for the classification of the soil type dataset. In the MobileNetV2 classifier, the edge and mobile devices are used with the integration of inverted residual and linear bottlenecks that ultimately create a deep neural network with low computational overhead. The depth-wise separable convolutions are used in a novel layer that ultimately reduces the total number of parameters, which enhances the efficacy of the model [33]. Furthermore, VGG16 deep learning architecture is also used in this work. It is a simple and uniform architecture that consists of 16 layers. This classifier also contains a series of convolutional layers of 3 × 3 filters and a max-pooling layer. This model is very efficient in extracting complex features [34]. At last, inception V3 architecture has asymmetric convolutions and factorized 7 × 7 convolutions, which ultimately enhances the utilization of computing resources. Furthermore, the most effective and reliable filer size for each convolution operation is chosen by hallmark inception modules. For feature extraction in the MobileNetV2 classifier, we use depth-wise separable convolution. The standard convolution is factorized into depth-wise convolution and 1 × 1 point-wise convolution, which helps in reducing the overall computational cost and total parameters and captures the most important features among them, as given in (2) [35].

$$Y = Y\text{pointwise} = \text{PointwiseConv}(\text{DepthwiseConv}(X)) \tag{2}$$

where, X represents the input tensor with dimensions (H, W, C). Here, H and W show the height and width of the input tensor, respectively, and C represents the total number of channels in the input tensor.

After that, we fine-tune the pre-trained model on a new data set having a low learning rate before using it for the classification of the soil type dataset. The weights are updated with a lower learning rate during fine-tuning, as given in (3). In this way, we ensure that the model has the generic features learned from the ImageNet dataset while adapting the most significant features of the soil images [33].

$$Wt + 1 = Wt - \alpha \cdot \nabla L(Wt) \tag{3}$$

where, Wt and Wt+1 represent the weights of the model at iteration t and updated weights of the model at iteration t+1, respectively. ∇L (Wt) is the gradient of loss function L with weight Wt. The learning rate is represented as α, which is set to the lower value in fine-tuning.

Secondly, we use stacked convolutional layers for feature extraction in the VGG16 classifier. There are stacked 3 × 3 convolutional layers and max-pooling layers in the VGG16 classifier. This structure helps in the extraction of the hierarchy of features, all the simple and complex patterns are extracted, which ultimately facilitates the soil type classification. Furthermore, we also use a fully connected layer at the end of the network for feature extraction. It provides a very distributed and high-level representation of input images. After that, we use the output of the last convolutional layer as bottleneck features. These features are then fed to the VGG16 classifier for training it for the classification of the soil type dataset. Lastly, we use inception modules for feature extraction for the InceptionV3 model.

The convolutions of different sizes are performed by inception modules due to which the model can capture features at various scales. In this way, the model can handle different patterns and textures in the images of soil. Furthermore, we fine-tune the inception V3 model by freezing earlier layers and retaining the latter layers. This helps in extracting the pre-trained features while simultaneously adapting specific properties of soil classification.

V. RESULTS AND DISCUSSION

This section includes results that are generated using the models on both datasets. There are 13,000 images in the dry beans' dataset, and we evaluate the performance of MLP, Naïve Bayes, and SVM classifiers in terms of F1-score, precision, recall, accuracy, and overall loss. We generate a confusion matrix, actual classes are given in rows and columns represent the classes, which are predicted by models. MLP model efficiently classifies the dry beans dataset. Table 1 shows the metrics for each class of beans using the MLP model. It can be observed from the table that the highest class predicted by the MLP model is 3rd class 'Cali' which accurately predicted 418 images. These results are also visible in the confusion matrix as shown in Fig. 9.

TABLE 1. Classification performance of MLP classifier.

Classes	Precision	Recall	F1-Score	Support
Barbunya	0.92	0.86	0.89	198
Bombay	1.00	0.64	0.78	78
Cali	0.83	0.96	0.89	245
Dermason	0.97	0.79	0.87	532
Horzo	0.98	0.93	0.95	289
Seker	0.99	0.92	0.95	304
Sira	0.71	0.96	0.82	396
	Accuracy: 0.88	Loss:0.35		

Table 1 shows that the MLP model shows a commendable classification performance with the dry beans dataset and achieves an accuracy of 0.88. It is also observed that the model successfully classifies 'Horzo' and 'Seker' beans with an F1-score of 0.95, which shows that this model identifies beans with minimal errors. However, there are some imbalances in the MLP model between precision and recall for 'Sira' and 'Bombay' classes. The model has a remarkable precision of 1 with 'Bombay' but it has a very small value of recall, which is 0.64, which shows that MLP misses a large number of 'Bombay' instances. Similarly, the value of recall for 'Sira' is 0.96 while it has a precision of 0.71, which shows that the model is misclassifying other beans as 'Sira'. The confusion matrix given in Fig. 9 shows the performance of the MLP model across all seven classes, the main diagonal represents correctly classified instances. The correct number of predictions is 418 for class 3, which means that the model has a very good prediction performance in identifying this class. While on the other hand, classes 1 and 6 have the highest number of false negative values. The reason is that these classes have some similar features with other classes as similarities between classes 3 and 6 are observed and the model is confused in identifying 6 and the same is the case with class 1. Overall, the performance of the MLP model is very high in the prediction of classes.

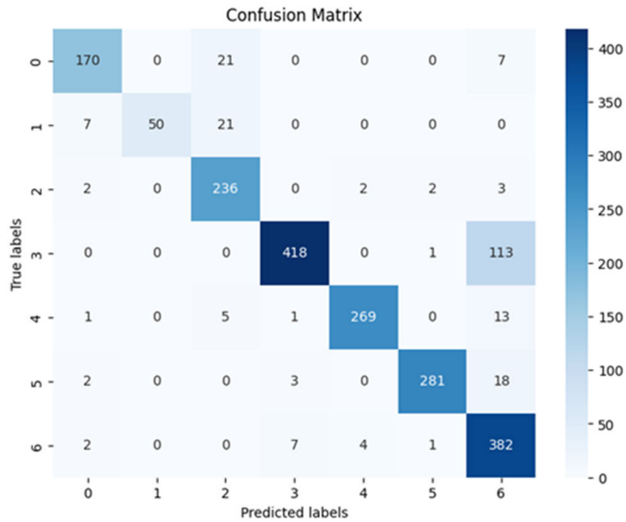

FIGURE 9. MLP classifier confusion matrix.

Table 2 represents the performance of the Naïve Bayes classifier on the dry beans' dataset. For the 'Bombay' class, the precision and recall of the classifier is 1, which represents that every instance in this class is correctly identified without any misclassification. This is due to the reason that the 'Bombay' beans class has less complex features. On the other hand, it is very challenging for the Naïve Bayes classifier to classify the 'Barbunya' class. The accuracy and recall of the Naïve Bayes classifier for this class are 0.68 and 0.47, respectively.

TABLE 2. Classification performance of naïve bayes.

Classes	Precision	Recall	F1-Score	Support
Barbunya	0.68	0.47	0.56	712
Bombay	1	1	1	682
Cali	0.63	0.78	0.70	676
Dermason	0.82	0.81	0.82	696
Horzo	0.77	0.77	0.77	756
Seker	0.76	0.73	0.75	724
Sira	0.69	0.78	0.73	719
	Accuracy: 0.77	Loss:0.30		

Furthermore, the F1-score is 0.56, which shows that some features of this class overlap with other classes, which ultimately leads to misclassification. The Naïve Bayes classifier shows moderate performance for the classification of 'Cali', 'Dermason', 'Horzo' and 'Sira' classes. The F1-score of this classifier for these classes is ranging from 0.70 to 0.82. the overall accuracy of the Naïve Bayes classifier is 0.77 with a loss of 0.30, which shows that the model is efficient in the classification of the dry beans' dataset.

Furthermore, the performance of the Naïve Bayes classifier is given in the confusion matrix as shown in Fig. 10. The result shows that the highest predicted class by this model was the 2nd class 'Bombay' which accurately predicted 682 images. It is also observed that class 1 shows no misclassification, which means that this class has distinctive features, that are well-captured by the model. The overall patterns of the confusion matrix indicate that the model can correctly and efficiently identify certain classes.

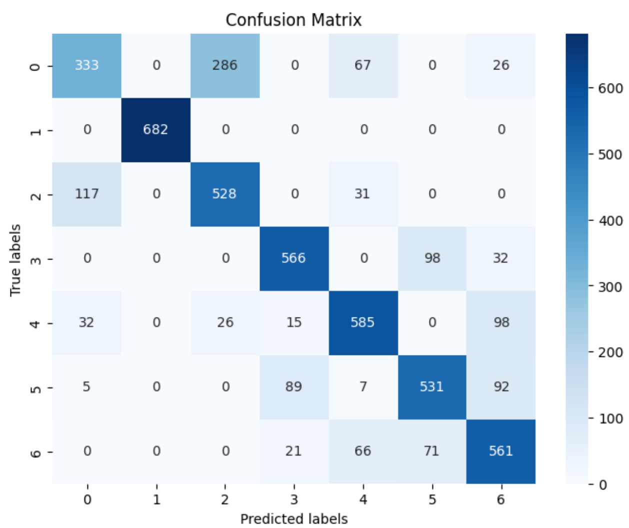

FIGURE 10. Naïve Bayes classifier confusion matrix.

Table 3 shows the performance of the SVM classifier for the dry beans' dataset. The overall accuracy of this classifier is 0.93. The precision and recall of this classifier for the 'Bombay' class is 1, which shows that there is no misclassification in this class. On the other hand, the precision and recall for 'Barbunya', 'Cali', 'Dermason', 'Horzo' and 'Seker' classes are very close, which means that SVM is capable of effectively differentiating the beans with high degrees of accuracy. The F1-score of the SVM classifier is consistent across diverse classes of beans, which means that this classifier made a robust decision boundary for this particular dataset.

TABLE 3. Classification performance of SVM classifier.

Classes	Precision	Recall	F1-Score	Support
Barbunya	0.95	0.88	0.91	429
Bombay	1	1	1	156
Cali	0.92	0.95	0.94	465
Dermason	0.92	0.93	0.92	1084
Horzo	0.96	0.96	0.96	555
Seker	0.95	0.96	0.95	591
Sira	0.87	0.87	0.87	804
	Accuracy: 0.93	Loss:0.21		

Furthermore, Fig. 11 shows the confusion matrix for the SVM classifier on the dry beans' dataset. There is a high positive rate for most of the classes. The highest predicted class by this model was the 4th class 'Dermason', which accurately predicted 1009 images.

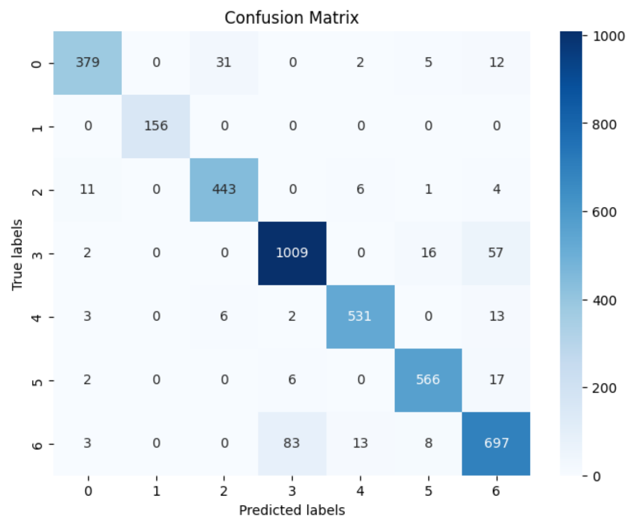

FIGURE 11. SVM classifier confusion matrix.

Table 4 represents the classification performance of a hybrid model of random forest and neural network on the dry beans' dataset. The results show that the proposed model has high precision and recall for all classes, which indicates that the hybrid model has effective learning and generalization capabilities. The results of this classifier for the 'Bombay' class show that it performs exceptionally well for this class, which shows that this 'Bombay' class has very high distinctive features and they are well captured by the classifier. Furthermore, the overall accuracy of the model is 0.92, which shows the effectiveness of the model. However, there is an improvement needed because the loss function of this classifier is 0.56.

TABLE 4. Classification performance of random forest and neural network classifier.

Classes	Precision	Recall	F1-Score	Support
Barbunya	0.94	0.91	0.93	198
Bombay	1	1	1	78
Cali	0.95	0.94	0.94	245
Dermason	0.90	0.92	0.91	532
Horzo	0.96	0.96	0.96	289
Seker	0.94	0.97	0.95	304
Sira	0.87	0.84	0.85	396
	Accuracy: 0.92	Loss:0.56		

Furthermore, Fig. 12 shows the confusion matrix for the hybrid classifier of random forest and neural network on the dry beans' dataset. The highest predicted class by this model was the 4th class 'Dermason' which accurately predicted 491 images. The results are shown in the confusion matrix. The overall results show that this hybrid model is very effective in classifying the dry beans dataset. The reason is that it leverages the interpretability of the random forest model and the learning depth of the neural network, which helps in effectively classifying the complex data of dry beans. Table 5 shows the performance of the hybrid SVM and neural network classifier on the dry beans' dataset. The model archives an overall accuracy of 0.92, which shows that this model is capable of identifying unique features of classes. Furthermore, the precision value is high for 'Barbunya' and 'Bombay classes' classes. Furthermore, the recall value of the model for 'Cali' and 'Seker' classes is high, which shows that this model is sensitive in the detection of these classes.

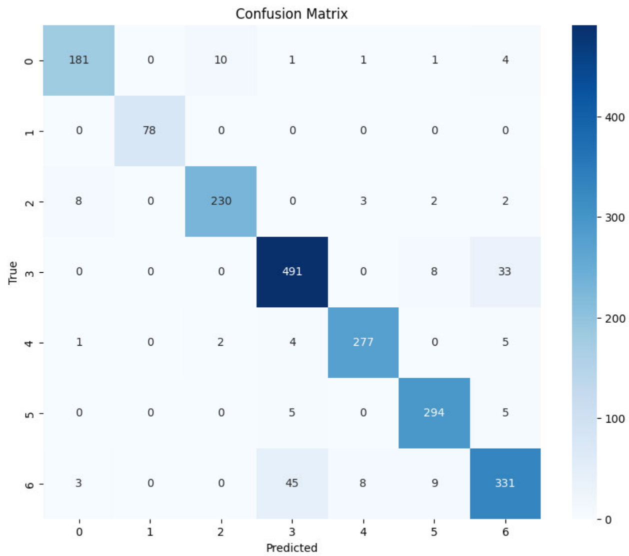

FIGURE 12. Random forest and neural network classifier confusion matrix.

TABLE 5. Classification performance of SVM and neural network classifier.

Classes	Precision	Recall	F1-Score	Support
Barbunya	0.95	0.90	0.93	198
Bombay	1	0.99	0.99	78
Cali	0.91	0.96	0.93	245
Dermason	0.90	0.91	0.91	532
Horzo	0.97	0.94	0.95	289
Seker	0.94	0.97	0.95	304
Sira	0.86	0.84	0.85	396
	Accuracy: 0.92		Loss:0.26	

Furthermore, the confusion matrix for the hybrid classifier of random forest and neural network on the dry beans' dataset is shown in Fig. 13. The highest predicted class by this model was also the 4th class 'Dermason' which accurately predicted 485 images. The results are shown in the confusion matrix. Moreover, classes 3 and 4 show the highest number of accurate predictions.

The soil type dataset contains 144 images, we evaluate the performance of the proposed model by considering the factors of accuracy, loss, and recall. Each model is trained on 30 Epochs and image classification took a longer time to train. Table 6 shows the classification performance of the MobileNetV2 model, which has an accuracy and recall of 0.97. It shows that the model can correctly identify the soil type around 97%. The high value of recall shows that the

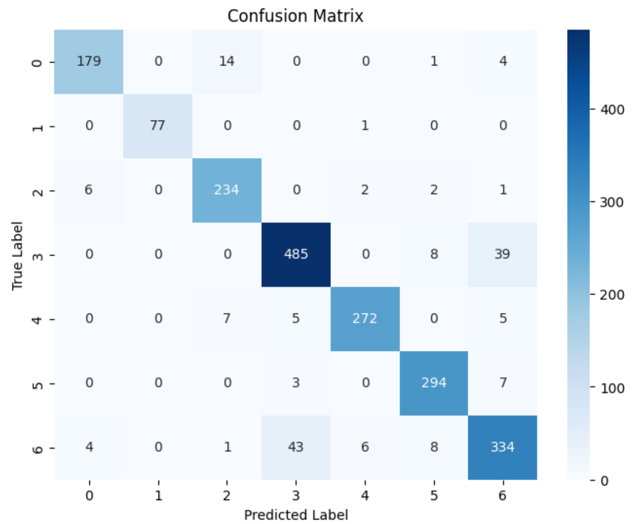

FIGURE 13. SVM and neural network classifier confusion matrix.

model performs well in identifying a high number of positive cases for each soil. On the other hand, the low loss rate of 0.08 indicates that the model can learn data with minimal errors, which makes it an efficient and reliable model for soil classification.

TABLE 6. Classification performance of MobileNetV2 classifier.

Accuracy	0.97
Loss	0.08
Recall	0.97

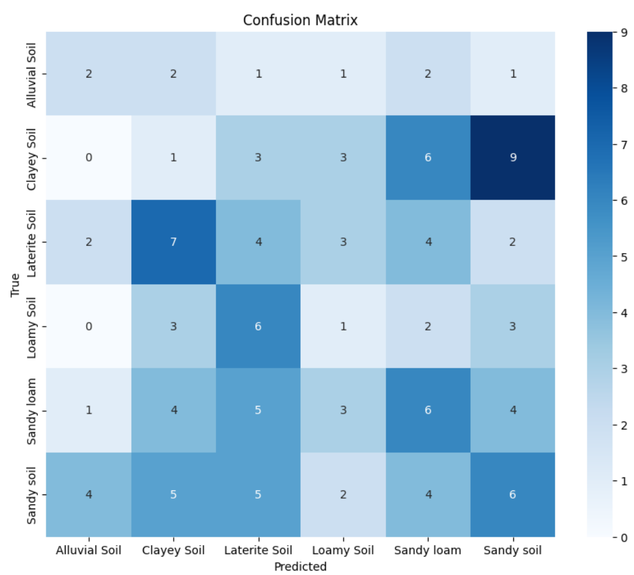

FIGURE 14. MobileNetV2 classifier confusion matrix.

The confusion matrix of soil classification with MobileNetV2 is given in Fig. 14. The results show that the highest correctly predicted class was the last class, 'Sandy

Soil'. The confusion matrix shows the strength of the model for classifying the soil dataset.

Table 7 shows that the VGG16 model has a high level of accuracy of 0.95 in classifying the soil type dataset, which means that the model is capable of correctly identifying 95% of instances. The value of recall is 0.94 of the VGG16 model, which shows that the model can correctly identify positive cases across different soil classes.

TABLE 7. Classification performance of VGG16 classifier.

Accuracy	0.95
Loss	0.20
Recall	0.94

Furthermore, the confusion matrix of soil classification with VGG16 is given in Fig. 15. The results show that the highest correctly predicted class is the 5th class, 'Sandy Loam'. The concentration of true positives in the prediction of 'Sandy Loam' is high because it has distinguished features and VGG16 can identify them effectively.

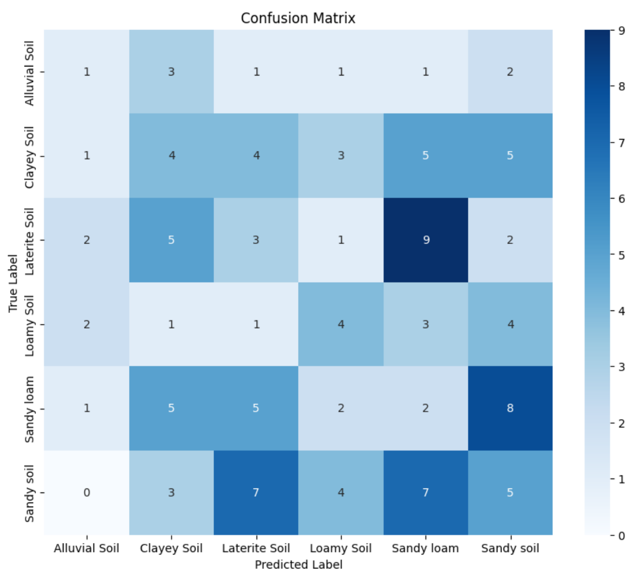

FIGURE 15. VGG16 classifier confusion matrix.

The classification performance of the InceptionV3 model is shown in Table 8. The overall accuracy of the model is 0.97, which indicates that the model is capable of correctly identifying a large majority of soil types. Furthermore, the recall of 0.95 shows that the model successfully captures 95% of positive instances in different classes of soil types. The loss of the model is 0.33, which shows that the model is not misclassifying a large number of instances.

Fig. 16 shows the confusion matrix for the performance of the inceptionV3 classifier. Certain types of soil are correctly identified with the proposed inception V3 model. It can be observed from the figure that the best-predicted class is also the 5th class, 'Sandy Loam' for this model as well. Table 9 shows the classification performance of the hybrid model of InceptionV3 with a Long Short-Term Memory (LSTM) network. The accuracy and recall of that hybrid model is 0.91, which indicates that the model is capable of identifying a high percentage of soil types. The reason is that the hybrid model uses the capabilities of InceptionV3'to to extract complex features while simultaneously using the strength of LSTM to process data sequentially.

TABLE 8. Classification performance of InceptionV3 classifier.

Accuracy	0.97
Loss	0.13
Recall	0.95

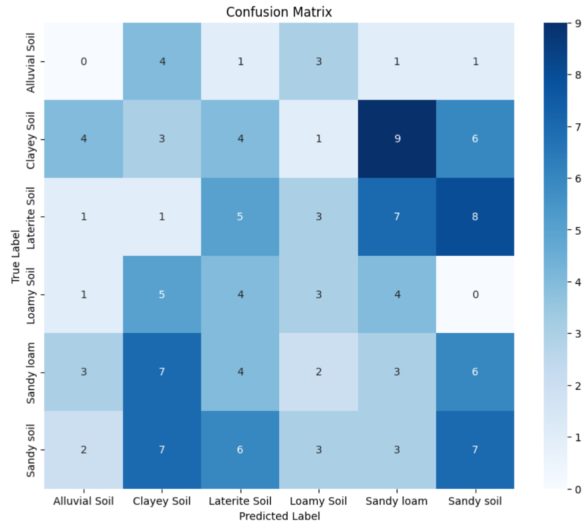

FIGURE 16. InceptionV3 classifier confusion matrix.

TABLE 9. Classification performance of InceptionV3 and LSTM classifier.

Accuracy	0.91
Loss	0.20
Recall	0.91

Furthermore, the confusion matrix of soil classification with the hybrid model of InceptionV3 and LSTM is given in Fig. 17. The results show that a high frequency of true positives is observed with the classification of 'Sandy Loam' and 'Sandy Soil'. Furthermore, the hybrid model has the same high score for many classes, which shows the classification effectiveness of the proposed hybrid model.

Table 10 shows the classification performance of the hybrid model of VGG16 and Dense layer Classifier. The results show that the accuracy of the hybrid model is 0.80. The reason is that the dense layer is augmented by the VGG16 model, and the model is then able to correctly classify 80% of the soil samples. Furthermore, the recall value of the model is 0.70, which means that the model is capable of capturing 70% of actual soil types. The reason is that the deep convolution

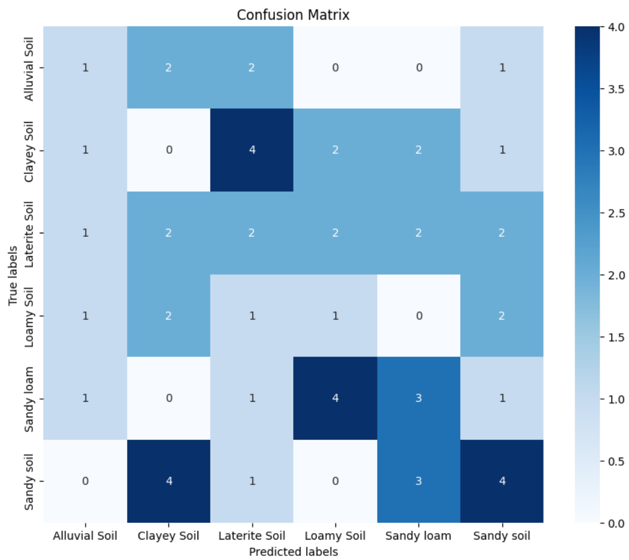

FIGURE 17. InceptionV3 and LSTM layer classifier confusion matrix.

layer of the VGG16 model is effective in extracting the most prominent features in the soil type dataset.

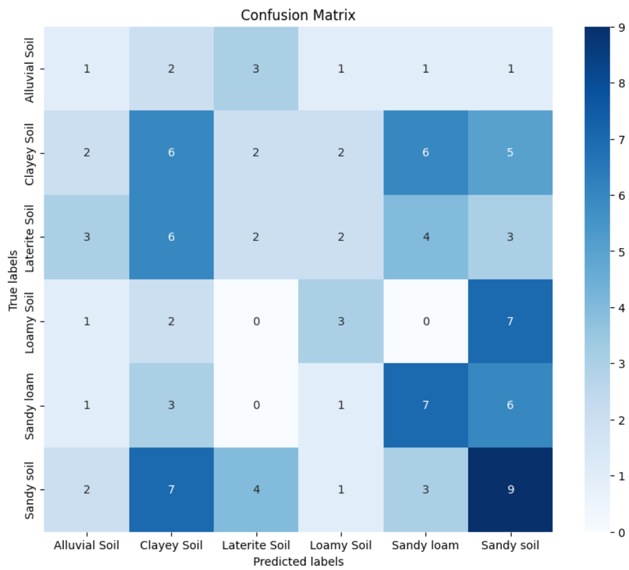

FIGURE 18. VGG16 and dense layer classifier confusion matrix.

TABLE 10. Classification performance of VGG16 and dense layer classifier.

Accuracy	0.80
Loss	0.60
Recall	0.70

In last, the confusion matrix of soil classification with the hybrid model of VGG16 and Dense layer is given in Fig. 18. The results show that a high frequency of true positives is observed with the classification of 'Sandy Loam' and 'Sandy Soil'. Furthermore, the correctly predicted class is the 6th class which is 'Sandy Soil'. Furthermore, the model has the same high score for many classes, which shows the classification effectiveness of the proposed hybrid model.

The bar chart given shows the comparative analysis of different ML algorithms for the classification of the dry bean dataset. It can be observed from Fig. 19 that the accuracy of the MLP algorithm is 80%, the reason is that this algorithm can learn very complex patterns, which makes it suitable for the classification of high dimensional and complex data. While the Naïve Bayes classifier has low accuracy, the reason is that this classifier assumes feature independence, which is not suitable for agricultural data where there is a high correlation between features. The SVM classifier, on the other hand, has around 91% accuracy because it is a robust model and effective for linearly separable data. The hybrid model of random forest and neural network has an accuracy of 92%. The reason is that random forest reduces overfitting and neural networks capture non-linear relationships while utilizing the properties of deep learning model. In last, a hybrid model of SVM and neural network also has a high accuracy of around 92%. The reason is that this model is also using the strength of individual classifiers. The SVM model defines the clear margin between classes and the neural network effectively learns the features and performs classification.

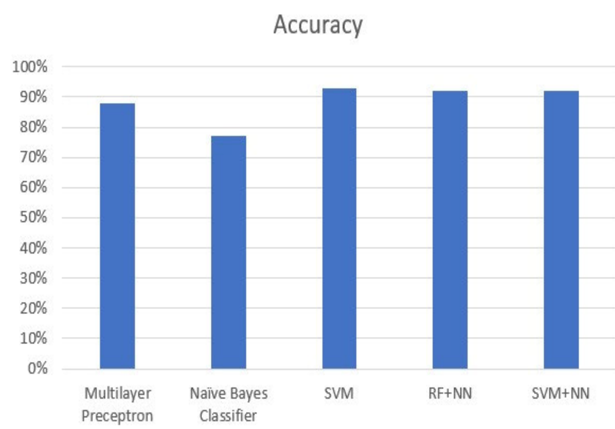

FIGURE 19. Comparative analysis of classifier accuracies for dry bean dataset.

Fig. 20 shows the comparative analysis of different deep-learning algorithms for the classification of soil-type datasets. It can be observed from this bar chart that the accuracy of the MobileNetV2 classifier is 85%, the reason is that this is a lightweight architecture that detects soil texture without large computational overhead. Similarly, the accuracy of the VGG16 model is around 84%, the reason is that this classifier is robust and effectively extracts features and nuance of the dataset. On the other hand, the accuracy of inceptionV3 is lower than the first two models. The reason is that the multi-scale features are not aligned with the dataset features.

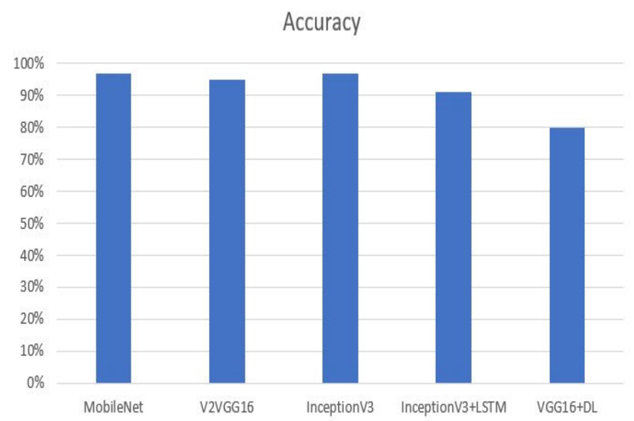

FIGURE 20. Comparative analysis of classifier accuracies for soil type dataset.

Furthermore, the hybrid model of inceptionV3 and LSTM has a slightly lower accuracy as compared to inceptionV3. This is due to the factor that the sequential pattern processing of LSTM does not help classify the dataset. In last, the hybrid model of VGG16 and the dense layer has the lowest accuracy among all. This may be due to the reason that there is overfitting due to the extra deep learning layer, which ultimately reduces the generalization of the model.

Fig. 21 shows the training and validation loss for the neural network over 20 epochs when it is applied to the dry beans dataset. The line graph shows that the training loss is very high in the initial stages at epoch 0, which means that the prediction of the model is not accurate at this stage. As with the increasing epochs, the loss of training decreases to a large extent, which indicates that the training of the model is improving. Around epoch 5, the training loss is reduced to 0.45. On the other hand, the validation loss also shows the same behavior as the training loss. It keeps on decreasing with the increasing epochs.

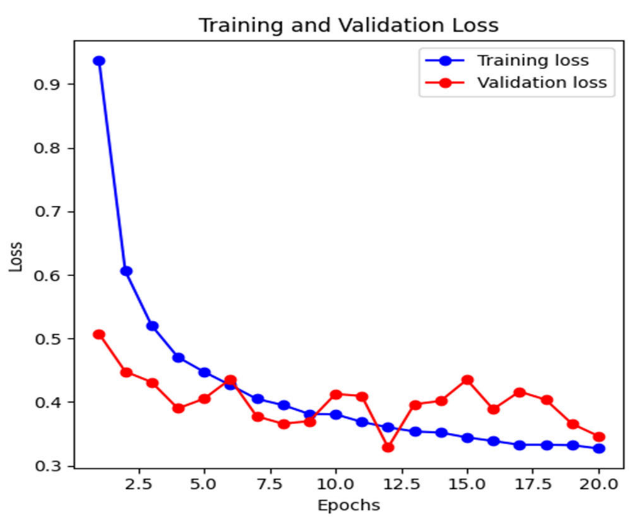

FIGURE 21. Validation and training loss of Neural Network.

Fig. 22 shows the training and validation accuracy of the neural network model when it is trained on the dry beans dataset over 20 epochs. Initially, the accuracy of training is very high. At the first epoch, it has a value of 0.86, which shows that the model can learn a large amount of portion of patterns in no time. This accuracy keeps on increasing with the increasing number of epochs; its value is 0.90 at epoch 5. Furthermore, the value of validation accuracy is 0.84 at the start, which means that its behavior is not consistent as compared to training data. It can also be observed that it has dips and upward trends with the passage of epochs.

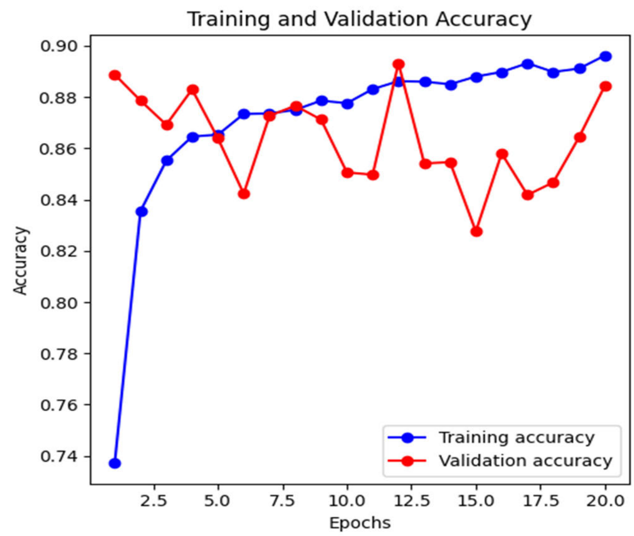

FIGURE 22. Validation and training accuracy of Neural Network.

Fig. 23 shows the values of training and validation loss for the hybrid model of random forest and neural network. The value of training loss decreases significantly from 0.8 to 0.1, which shows that this model has the capabilities of quick learning and is a strong fit for the training data. Furthermore, the value of validation loss is 0.65 at the start and it keeps on increasing after a slight decrease, which shows that the model can learn training data efficiently.

Similarly, Fig. 24 shows the training and validation accuracy of hybrid models of random forest and neural networks when this model is applied to the dry beans dataset. It can be observed that the training accuracy increases to a large extent at epoch 2 and then keeps this behavior throughout, which means that the model can excellent, fit to the training data. On the other hand, the validation accuracy is much lower as compared to the training accuracy, which shows that the model is not able to show generalization to unseen data.

The training and validation losses for the hybrid model of support vector machine and neural network are shown in Fig. 25. At the start, the training loss is very high, having a value of around 0.6. After that, it sharply decreases to a large extent by the third epoch, which shows that the model can quickly learn early in the training process. Moreover, it can

FIGURE 23. Validation and training loss of random forest + neural network.

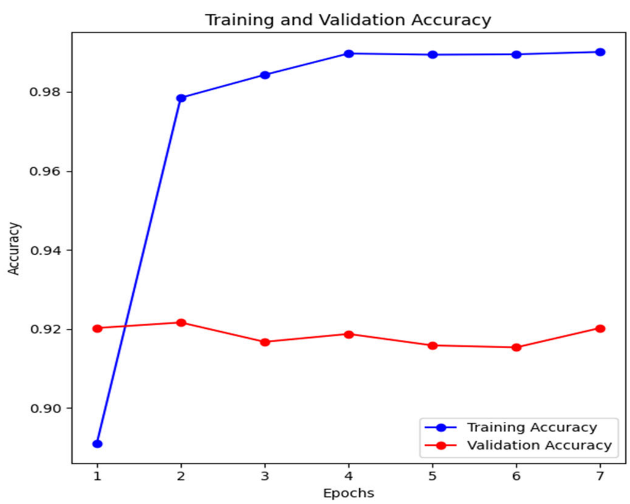

FIGURE 24. Validation and training accuracy of random forest + neural network.

also be observed that both the training and validation losses decrease and begin to converge at around the 10th epoch.

Fig. 26 shows the training and validation accuracy of hybrid models of support vector machine and neural network. The accuracy of training starts around 90% and quickly increases to 92% at the second epoch. On the other hand, the validation accuracy aligns closely with the training accuracy and stays above 90%, which shows that the model has the capability of generalization without substantial overfitting.

VI. LIMITATIONS OF PROPOSED

The proposed model is efficient for the classification of data with dry beans and soil-type datasets. However, the ability of the model to generalize to other agricultural datasets or different conditions is not evaluated in this study. Furthermore, it is very challenging to implement complex hybrid models in

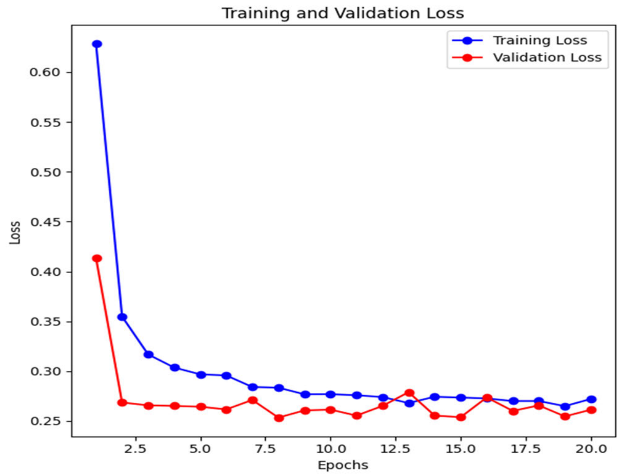

FIGURE 25. Validation and training loss of SVM + neural network.

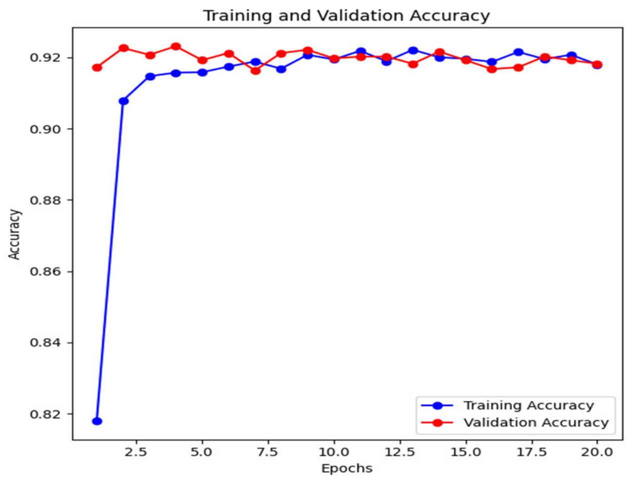

FIGURE 26. Validation and training accuracy of SVM + neural networks.

real-world agricultural environments such as scalability and integration with existing farming systems. In last, there is no security measure in the IoT-enabled framework to tackle the issues of single point of failure, replay attack, denial of service attack, and distributed denial of service attack.

VII. CONCLUSION

In this paper, we introduce an effective and robust IoT-enabled hybrid ML and deep learning model for smart agriculture, for efficiently classifying the dry beans and soil-type datasets. By effectively combining traditional ML algorithms with advanced deep learning techniques, our models demonstrated significant potential in accurately classifying dry bean classes and differentiating soil types, with the best models achieving accuracies up to 97%. Despite these promising results, we acknowledge the need for further optimization to address computational efficiency and model complexity. Specifically, the hybrid model featuring the traditional SVM classifier and novel neural network achieved an exceptional accuracy of 0.92, showcasing its efficacy in identifying dry bean classes. Furthermore, in the case of the soil type dataset, MobileNetV2 outperformed all other models with an exceptional accuracy of 0.97, which showed the effectiveness of deep learning in differentiating complex soil patterns. Besides this, we also proposed a novel hybrid technique by combining InceptionV3 with LSTM networks, which yielded an accuracy of 0.91. This underscores the model's adeptness in handling sequential and spatially complex data. Lastly, the hybrid model combining random forest and neural network achieved an accuracy of 92%, further validating the effectiveness of our approach. In the future, we will enhance model abilities by proposing more efficient and reliable data augmentation techniques. Furthermore, an authentication mechanism will be proposed for IoTs network for data access control, which ultimately helps in ensuring the integrity of data in the network.

ACKNOWLEDGMENT

The authors would like to thank the Deanship of Scientific Research, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia, for providing research resources and equipment.

REFERENCES

[1] I. Attri, L. K. Awasthi, and T. P. Sharma, "Machine learning in agriculture: A review of crop management applications," *Multimedia Tools Appl.*, vol. 83, no. 5, pp. 12875–12915, Jul. 2023.

[2] M. Raj, S. Gupta, V. Chamola, A. Elhence, T. Garg, M. Atiquzzaman, and D. Niyato, "A survey on the role of Internet of Things for adopting and promoting agriculture 4.0," *J. Netw. Comput. Appl.*, vol. 187, Aug. 2021, Art. no. 103107.

[3] B. B. Sinha and R. Dhanalakshmi, "Recent advancements and challenges of Internet of Things in smart agriculture: A survey," *Future Gener. Comput. Syst.*, vol. 126, pp. 169–184, Jan. 2022.

[4] R. Akhter and S. A. Sofi, "Precision agriculture using IoT data analytics and ML," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 8, pp. 5602–5618, 2022.

[5] O. Friha, M. A. Ferrag, L. Shu, L. Maglaras, and X. Wang, "Internet of Things for the future of smart agriculture: A comprehensive survey of emerging technologies," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 4, pp. 718–752, Apr. 2021.

[6] B. J. Chelliah, T. P. Latchoumi, and A. Senthilselvi, "Analysis of demand forecasting of agriculture using machine learning algorithm," *Environ., Develop. Sustainability*, vol. 26, no. 1, pp. 1731–1747, Dec. 2022.

[7] M. F. Guerri, C. Distante, P. Spagnolo, F. Bougourzi, and A. Taleb-Ahmed, "Deep learning techniques for hyperspectral image analysis in agriculture: A review," *ISPRS Open J. Photogramm. Remote Sens.*, vol. 12, Apr. 2024, Art. no. 100062.

[8] R. R. Chaudhary, S. Jain, and S. Gupta, "A critical review on hybrid framework for precise farming with application of machine learning (ML) and Internet of Things (IoT)," *J. Integr. Sci. Technol.*, vol. 12, no. 2, p. 730, 2024.

[9] K. Phasinam, T. Kassanuk, and M. Shabaz, "Applicability of Internet of Things in smart farming," *J. Food Qual.*, vol. 2022, pp. 1–7, Feb. 2022.

[10] A. D. Boursianis, M. S. Papadopoulou, P. Diamantoulakis, A. Liopa-Tsakalidi, P. Barouchas, G. Salahas, G. Karagiannidis, S. Wan, and S. K. Goudos, "Internet of Things (IoT) and agricultural unmanned aerial vehicles (UAVs) in smart farming: A comprehensive review," *Internet Things*, vol. 18, May 2022, Art. no. 100187.

[11] N. Huynh and K.-D. Nguyen, "Real-time droplet detection for agricultural spraying systems: A deep learning approach," *Mach. Learn. Knowl. Extraction*, vol. 6, no. 1, pp. 259–282, Jan. 2024.

[12] J. Rui and S. Danpeng, "Architecture design of the Internet of Things based on cloud computing," in *Proc. 7th Int. Conf. Measuring Technol. Mechatronics Autom.*, Jun. 2015, pp. 206–209.

[13] O. Friha, M. A. Ferrag, L. Shu, L. Maglaras, K.-K.-R. Choo, and M. Nafaa, "FELIDS: Federated learning-based intrusion detection system for agricultural Internet of Things," *J. Parallel Distrib. Comput.*, vol. 165, pp. 17–31, Jul. 2022.

[14] W. Tao, L. Zhao, G. Wang, and R. Liang, "Review of the Internet of Things communication technologies in smart agriculture and challenges," *Comput. Electron. Agricult.*, vol. 189, Oct. 2021, Art. no. 106352.

[15] S. Pasika and S. T. Gandla, "Smart water quality monitoring system with cost-effective using IoT," *Heliyon*, vol. 6, no. 7, Jul. 2020, Art. no. e04096.

[16] M. V. Ramesh, K. V. Nibi, A. Kurup, R. Mohan, A. Aiswarya, A. Arsha, and P. R. Sarang, "Water quality monitoring and waste management using IoT," in *Proc. IEEE Global Humanitarian Technol. Conf. (GHTC)*, Oct. 2017, pp. 1–7.

[17] M. Kuradusenge, E. Hitimana, K. Mtonga, A. Gatera, J. Habiyaremye, J. Ngabonziza, D. Hanyurwimfura, P. Rukundo, and A. Mukasine, "SMART-CYPS: An intelligent Internet of Things and ML powered crop yield prediction system for food security," *Res. Square*, vol. 1, no. 1, pp. 1–22, Mar. 2024, doi 10.21203/rs.3.rs-3834903/v1.

[18] Y.-H. Tsai and T.-C. Hsu, "An effective deep neural network in edge computing enabled Internet of Things for plant diseases monitoring," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops (WACVW)*, Jan. 2024, pp. 695–699.

[19] M. R. Machado, T. R. Júnior, M. R. Silva, and J. B. Martins, "Smart water management system using the microcontroller ZR16S08 as IoT solution," in *Proc. IEEE 10th Latin Amer. Symp. Circuits Syst. (LASCAS)*, Feb. 2019, pp. 169–172.

[20] Y. Khan and C. S. See, "Predicting and analyzing water quality using machine learning: A comprehensive model," in *Proc. IEEE Long Island Syst., Appl. Technol. Conf. (LISAT)*, Apr. 2016, pp. 1–6.

[21] A. Gupta and P. Nahar, "Classification and yield prediction in smart agriculture system using IoT," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 8, pp. 10235–10244, Aug. 2023.

[22] A. Chauhan and P. Tripathy, "Internet of Things (IoT) integrated solutions for environmentally friendly intelligent farming: A systematic review," in *Proc. 3rd Int. Conf. Advance Comput. Innov. Technol. Eng. (ICACITE)*, May 2023, pp. 2118–2123.

[23] A. Vangala, A. K. Das, and J. Lee, "Provably secure signature-based anonymous user authentication protocol in an Internet of Things-enabled intelligent precision agricultural environment," *Concurrency Comput., Pract. Exper.*, vol. 35, no. 16, p. e6187, Jul. 2023.

[24] C. Dang, Y. Liu, H. Yue, J. Qian, and R. Zhu, "Autumn crop yield prediction using data-driven approaches: Support vector machines, random forest, and deep neural network methods," *Can. J. Remote Sens.*, vol. 47, no. 2, pp. 162–181, Mar. 2021.

[25] T. Van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using ML: A systematic literature review," *Comput. Electron. Agricult.*, vol. 177, no. 2020, Oct. 2020, Art. no. 105709.

[26] S. H. Bhojani and N. Bhatt, "Wheat crop yield prediction using new activation functions in neural network," *Neural Comput. Appl.*, vol. 32, no. 17, pp. 13941–13951, Sep. 2020.

[27] Dry Bean Dataset. (2021). *Kaggle Repository*. [Online]. Available: https://www.kaggle.com/datasets/sansuthi/dry-bean-dataset/data

[28] M. A. Imron, M. Ali, and B. Prasetyo, "Improving algorithm accuracy K-nearest neighbor using Z-score normalization and particle swarm optimization to predict customer churn," *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 56–62, 2020.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[30] W. E. Finch-Savage, "Influence of seed quality on crop establishment, growth, and yield," in *Seed Quality*. Boca Raton, FL, USA: CRC Press, 2020, pp. 361–384.

[31] F. L. Gewers, G. R. Ferreira, H. F. De Arruda, F. N. Silva, C. H. Comin, D. R. Amancio, and L. da F. Costa, "Principal component analysis: A natural approach to data exploration," *ACM Comput. Surv.*, vol. 54, no. 4, pp. 1–34, 2021.

[32] G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," 2013, *arXiv:1302.4964*.

[33] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[34] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[35] B. Karsh, R. H. Laskar, and R. K. Karsh, "MIV3Net: Modified inception V3 network for hand gesture recognition," *Multimedia Tools Appl.*, vol. 83, no. 4, pp. 10587–10613, Jan. 2024.

MOHAMMAD ALDOSSARY received the B.Sc. degree (Hons.) in computer science from King Saud University, Riyadh, Saudi Arabia, in 2009, the M.Sc. degree in computer science from Southern Polytechnic State University, GA, USA, in 2013, and the Ph.D. degree in computer science from the University of Leeds, U.K., in 2019. He is currently an Associate Professor with the Department of Computer Engineering and Information, Faculty of Engineering College, Prince Sattam Bin Abdulaziz University (PSAU), Saudi Arabia. His main research interests include distributed systems, including cloud, fog, and edge computing and the Internet of Things (IoT) technologies, such as the Internet of Medical Things (IoMT) and the Industrial Internet of Things (IIoT), artificial intelligence (AI), machine learning (ML), digital twin, metaverse, smart cities, unmanned aerial vehicles (UAVs), smart agriculture, system architectures, resource management, and energy efficiency. He holds the honor of being a Founding Member of the Artificial Intelligence Governance Association (AIGA), Saudi Arabia; and serves as a Board Member for the Saudi Internet of Things Association, Saudi Arabia.

HATEM A. ALHARBI received the B.Sc. degree (Hons.) in computer engineering from Umm Al-Qura University, Makkah, Saudi Arabia, in 2012, and the M.Sc. degree (Hons.) in digital communication networks and the Ph.D. degree in communication networks from the University of Leeds, Leeds, U.K., in 2015 and 2020, respectively. He is currently an Assistant Professor with the Computer Engineering Department, School of Computer Science and Engineering, Taibah University, Saudi Arabia. His research interests include energy efficient fog and cloud networks and network economics.

CH ANWAR UL HASSAN received the B.S. degree in software engineering from the National University of Modern Languages, Islamabad, Pakistan, in 2015, and the M.S. degree in software engineering from COMSATS University, Islamabad, in 2019. After graduation, he worked as a Visiting Lecturer at the Federal Urdu University of Arts, Science and Technology, Islamabad. He also worked as a Lecturer at the Capital University of Science and Technology, Islamabad. Currently, he is a Lecturer at Air University, Islamabad. He is an experienced Software Engineer, having worked as a Web/App Developer from 2015 to 2017, where he was involved in the development of various top-notch software projects, participating in all stages from documentation to deployment. From 2017 to 2019, he was a Research Assistant with COMSATS University. His research interests include blockchain, data warehousing, data analysis, machine learning, data mining, smart grid energy management, software process improvements, and software costing and estimation.