energies

MDPI

Article

Smart Agriculture Cloud Using AI Based Techniques

Muhammad Junaid [1,*], Asadullah Shaikh [2,*], Mahmood Ul Hassan [2], Abdullah Alghamdi [2], Khairan Rajab [2,*], Mana Saleh Al Reshan [2] and Monagi Alkinani [3]

[1] Department of Information Technology, The University of Haripur, Haripur 22620, KPK, Pakistan
[2] College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia; mahmood.mscs@gmail.com (M.U.H.); abdulresearch@hotmail.com (A.A.); msalreshan@nu.edu.sa (M.S.A.R.)
[3] Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering, University of Jeddah, Jeddah 21442, Saudi Arabia; malkinani@uj.edu.sa
* Correspondence: mjunaid@uoh.edu.pk (M.J.); asshaikh@nu.edu.sa (A.S.); khairanr@gmail.com (K.R.)

Abstract: This research proposes a generic smart cloud-based system in order to accommodate multiple scenarios where agriculture farms using Internet of Things (IoTs) need to be monitored remotely. The real-time and stored data are analyzed by specialists and farmers. The cloud acts as a central digital data store where information is collected from diverse sources in huge volumes and variety, such as audio, video, image, text, and digital maps. Artificial Intelligence (AI) based machine learning models such as Support Vector Machine (SVM), which is one of many classification types, are used to accurately classify the data. The classified data are assigned to the virtual machines where these data are processed and finally available to the end-users via underlying datacenters. This processed form of digital information is then used by the farmers to improve their farming skills and to update them as pre-disaster recovery for smart agri-food. Furthermore, it will provide general and specific information about international markets relating to their crops. This proposed system discovers the feasibility of the developed digital agri-farm using IoT-based cloud and provides solutions to problems. Overall, the approach works well and achieved performance efficiency in terms of execution time by 14%, throughput time by 5%, overhead time by 9%, and energy efficiency by 13.2% in the presence of competing smart farming baselines.

Keywords: smart farming; AI-based agri-food; energy efficiency; digital transformation; environment; cloud based IoTs

Citation: Junaid, M.; Shaikh, A.; Hassan, M.U.; Alghamdi, A.; Rajab, K.; Al Reshan, M.S.; Alkinani, M. Smart Agriculture Cloud Using AI Based Techniques. *Energies* **2021**, *14*, 5129. https://doi.org/10.3390/en14165129

Academic Editors: Adnan Akhunzada, Philippe Roose and André Fonteles

Received: 14 July 2021
Accepted: 11 August 2021
Published: 19 August 2021

1. Introduction

The introduction should briefly place the study in a broad context and highlight why it is important. Food requirements are heavily dependent on agriculture and its quality of production. There is a large number of countries in the world where agriculture makes a large contribution to their economies. The populous countries such as Pakistan, India, China, etc., have a major portion of their land devoted to agriculture which not only fulfills their own needs but also provide handsome foreign return [1]. Currently, the integration of agriculture with internet technologies has contributed value addition to traditional farming. Similarly, the advent of the Internet of Things (IoT) in the agriculture sector has changed traditional farming to smart farming. Using cloud computing service models for storing and accessing agriculture data from multiple sources has greatly impacted the performance in the offline and real-time environment [2,3]. This performance in the cloud is measured by multiple factors such as response time, execution time, overhead time, migration time, optimization time, and energy consumption by computing resources [4]. Similarly, smart farming, with the help of IoTs, is using powerful machine learning models to access and disseminate the data in an efficient manner from and to the end-users such as farmers, advisors, and researchers. There is a strong need to process real-time information accurately so that an appropriate decision can be made in a timely manner by the end-users. This is only possible with the inclusion of strong machine learning models, such as the Support Vector Machine (SVM), in the cloud environment for efficient information processing. Some machine learning algorithms such as Cat Swarm Optimization (CSO) are more suitable for small population sizes with a minimum number of iterations and, hence, do not provide good solutions in a situation where the processing involves a large number of complex tasks. This drawback eventually leads CSO to fall into local optimum, which takes more iterations in finding solution spaces and, hence, renders CSO computationally complex. Therefore, we used SVM by introducing a new grouping phase process that takes the data files into four groups (audio, video, image, and text) taken from the SVM, keeping in view the properties associated with each group.

Cloud computing is a type of internet computing in which users may access programs via a browser while the application (together with the data) is installed and kept on a server. This is a whole new type of computing that allows thousands of farmers all over the world to access information without having to download or install anything on their computers or mobile phones. It enables mobility in the same way that most networks allow users to connect even if they do not have their computers, allowing them to work from anywhere in the world as long as they have an internet connection and access to a computer. Sensor technologies, personal mobile devices, wireless broadband connections, and cloud computing are allowing agriculturists and regular farmers to collect and disseminate agricultural data in real-time from anywhere. Personal mobile devices, such as Personal Digital Assistant (PDA)s and cellphones, are becoming more capable in terms of processing and information management, and they play an increasingly important part in people's everyday lives. This technical progress has led us to develop a scalable and cost-effective real-time agricultural monitoring and analysis system for those who need to monitor their farms regularly [5]. The focus will be on the design features of an autonomous cloud environment that collects farmer's agricultural data and distributes it to a cloud-based information repository, as well as facilitating data analysis by utilizing cloud-based software applications.

The consideration of this research work is to create a cloud computing infrastructure that will connect different agriculture resources to a central data store. The services may be in the form of information/responses to different queries raised by the farmers; automatic suggestions to farmers from time to time to improve their farming skills; provide specific slots in lands where the problem exists in multiple formats; automatically update the farmers on which crop is suitable in their area based on various factors; provide special features to farmers where they can also interact with local and world-leading agriculturists; update the farmers as pre-disaster recovery; and provide general and specific information about international markets relating to their crops. The data collected by the concerned providers will be sent to the data cloud store via an underlying computerized system, where it will be accessible to all connected farmers, advisors, researchers, and support centers. There are several activities to be performed by the developed system such as acquiring, managing, and using agricultural information from farms to enhance the quality and efficiency of farms and to respond to widespread public farmers queries. In particular, at the automated farm level, agriculture envisions that "a new system of distributed computing tools will collect authorized data about farm illness and store it securely within a network designed to help deliver quick and efficient care". Furthermore, the concepts, such as wearable agriculture devices, Farm Area Networks (FANs), pervasive wireless broadband communications, and cloud computing, are enabling advanced mobile farm-care services that benefit both farmers and agriculture professionals [6]. This enables the development of a system to perform remote real-time collection, dissemination, and analysis of farm data to manage chronic conditions and to detect farm emergencies.

Our goal is to offer an architecturally generic cloud-based system that can be used in a variety of scenarios where farms need to be remotely monitored and recorded data have to be processed by a computer system and made available for professionals or farmers to access. Despite the fact that our designs and prototypes are flexible enough to suit a variety of use cases, we focus on one particular motivated case: The monitoring of farms affected by chemicals and sickness, which necessitates constant episode detection. Electrocardiogram data from commonly available wearable sensors are collected in real-time and utilized to detect and classify episodes.

This research study develops crop stress indicator models that provide data filter and alert capabilities for monitoring local agricultural conditions and provides an immediate solution to farmers as a remedy. Secondly, by taking the information service demand of "Three Rural Issues" in our country into consideration, we propose a cloud service platform for the information service in the countryside which can enable data sharing, remote data storage, interaction with farmers, agriculture expert consultation, and peasant household management. We want to provide a platform for the farmers with the answers to the following generalized question: How do our agricultural activities increase the yield or deliver higher efficiency with quality? As agriculture goes high-tech, our proposed agricultural technology revolution will not require farmers to put down their hoes and pick up a mouse. Instead, we predict the growth of specialist project groups with expertise in agricultural and associated disciplines, such as ecosystems [7]. Those who demand continual monitoring but live distantly from their service provider and find it impossible to attend frequent pesticides sessions will appreciate the value of a ubiquitous agricultural information system. Accordingly, it has been declared that the implemented approach has no adverse effect on the natural environment, social security, damage to human health, safety to biodiversity, pollution to physical, biological, hydrological resources, damage to public comfort, vegetal canopy, and other adverse environmental effects.

The following are the contributions of this research:

- Agricultural data received in the cloud are classified into different file formats such as audio, video, text, images, and maps files;
- File format classification is an innovative approach in smart farming that has never been adopted and, hence, contributed to the theory of science;
- Multi-format files are processed by using SVM one-to-many types of machine learning methods in order to achieve efficiency in the cloud;
- An AGRICLOUD (Agriculture Cloud) approach has achieved maximum accuracy with less response time, low overhead, and high throughput in the presence of a certain state of the art;
- Importantly, the farmer achieves the supervisory level so that decisions can easily be taken by him.

The rest of this research study is organized as follows: A literature review is presented in Section 2, the proposed methodology is discussed in Section 3, results are explored in Section 4, and discussions are elaborated in Section 5. Finally, Section 6 concludes the work and identifies the future directions.

2. Materials

Literature Review

Cloud computing is a new paradigm of computing in which dynamically scalable and often virtualized resources are referred to as clouds and are provided as a service over the internet [8]. Users of these services do not need any knowledge to construct them, maintain them, or to control the technology infrastructure used to construct these "clouds" that support them. Cloud computing technology is forcing a paradigm shift from owing the information and communication infrastructure to leasing it, allowing "cloud systems" to penetrate the market as an infrastructure in which services can be orchestrated. Cloud computing has emerged as a novel way for large-scale aggregation of various IT services provided through fast digital networks, similar to the electrical grid of public utilities [9,10].

The rise of cloud computing is rapidly changing the landscape of information technology and ultimately turning the long-held promise of utility computing into a reality [11,12].

Cloud computing's emergence has had a significant impact on the Information Technology (IT) industry in recent years, with companies such as Google, Amazon, and Microsoft contending to provide more powerful, reliable, and cost-effective cloud platforms. More businesses are reshaping their business models to take advantage of this evolving field [13,14]. There are several advantages of the cloud computing paradigm over traditional IT services on the following basis.

Minimum infrastructure investment cost: A cloud service user can simply rent the resources from different clouds, and the users only pays for the resources they have used according to some pricing model.

Minimum operating cost: A cloud service provider can dynamically allocate and de-allocate resources according to system usage situation at peak loads, which allows significant savings relative to the operational cost of resources when the usage of the system is low.

Scalability: A cloud service provider can easily extend its services by provisioning a larger amount of resources and making them accessible. The current IT technologies are not matured enough to realize the full potential of cloud computing technology [15]. Major challenges that arise in the implementation of cloud computing technology, including automatic resource provisioning, security management, and power management, are recieving attention from the research community. Therefore, researchers in this area still have tremendous opportunities to make novel contributions and bring about significant impact with respect to IT Industry.

Keeping in view the importance of cloud computing in the agriculture sector, several innovations have made inevitable the transition from traditional farming to smart farming in the world. Developing countries, such as Pakistan, need massive efforts to uplift their major source of revenue in the form of smart agriculture. To meet the food needs of Pakistan, continued population growth in Pakistan and the growing reliance on agriculture to provide replacements for fossil fuels mean that there are no alternatives but to begin smart and high-tech farming; all this simply amounts to using cloud computing to improve agricultural reforms.

Subsequently, farmers in rural and remote areas do not have similar access to intensive specialist support as in urban areas. Thus, there is a need to develop such a system that provides farmers with equal access to specialist support. Higher temperatures, increased agricultural water demand, more variable rainfall, and extreme climatic events such as heatwaves, floods, and droughts all have an impact on agriculture [16]. Farmers are the most vulnerable to climate change, but they can also play an important part in combating it by seizing opportunities.

Our proposed solution will investigate the above-mentioned problems for the benefit of the community at large. It would assist the farmers by providing a variety of possible options and also recommending the best solution based on many factors stored in a cloud of agricultural records in a timely and more efficient manner. The agricultural record includes a variety of types of "notes" entered over time by agriculturists and professionals from all over the world, including recording observations and the administration of pesticides, orders for the drugs, test results, scanning farms, reports, futuristic issues after natural disasters to lands, etc. The maintenance of a complete and accurate agricultural record is a fundamental requirement for any agriculturist and is generally enforced as a licensing or certification prerequisite.

Cloud Computing technology provides a pervasive means of efficiently sharing educational material and knowledge among institutions in a relatively low-cost manner in order to leverage the cooperative and sharing culture of higher education [17]. Collaboration between researchers, students, degree programs, and even administrative services is both a requirement and a driver for a true knowledge-based economy [18]. Cloud computing enables cost-effective collaboration across research-intensive institutions throughout the country and around the world, especially in the sciences, where large-scale equipment has become the standard. Scientists from diverse universities can share supercomputers, libraries can share digital humanities collections, astronomers can share galactic pictures, and network engineers may share strands of fiber in the same physical cable thanks to cloud computing technologies. In the same way, cloud computing is making a shifting paradigm by strengthening its roots in the agriculture sector. By establishing collaboration of scientists all over the world in the diverse field of agriculture, the end users can obtain maximum advantage in solving the number of problems with efficient solutions. This is only possible with the adoption of a cloud that relies not only on scientists but also takes into account other stakeholders such as farmers, advisors, local agriculture counselors, and students researching the same domain.

The current research study in smart agriculture is more focused on processing relevant information in a much shorter period of time. This includes the shortest response time in accessing the relevant information, the low overhead of the communication, lightweight protocol development for the execution time, and secure transmission of the critical information. To cater to such multiple QoS mentioned above, cloud computing is taking advantage of various machine learning models due to their established authenticity in many applications. One of the widely used machine learning models is SVM for making accurate classifications of data files. In the proposed agriculture model, the cloud would have a huge volume and variety of data taken from multiple sources such as IoTs, Zigbee, Wireless Sensor Network (WSN)s, etc. These data are securely stored in the cloud and will be updated continuously upon receiving information from farmers, scientists, advisors, and researchers. A variety of data in multi-formats such as audio, video, image, text, and maps are taken in real-time and stored in the cloud where SVM is used to classify them in respective classes. This would help in the accurate classification of the data files, which will be fed into Virtual Machines (VMs) for processing. In order to further improve efficient processing, one-to-many classification types of SVM are used. The process would be made radially available for the end-users for a prompt decision about multiple cases. An SVM model represents training data as points in space which are then separated into categories by a wider gap. The classifier then finds the hyperplane that differentiates two classes, and then it is categorized according to the gap they belong to. Furthermore, SVMs are effective in high dimensional spaces and in instances where the amount of dimensions is larger than the number of samples, and it is also memory efficient because it uses a subset of training data in the decision function which are called the support vectors. SVM treats one class at a time (e.g., text class) as a positive class and three other classes (audio, video, and image) as negative classes. Therefore, other classes are taken separately, in the same manner, using one-to-many classification properties of SVM. Some recent research studies taken from the literature are listed in Table 1, with some detail of their technique, their advantages, and disadvantages. Our designed method tries to minimize the limitations of the following research and proposes a comprehensive solution for an efficient agriculture cloud. In addition to the multi-format data classification approach using SVM, the proposed solution would curtail execution time in executing the query and generating the response to be received by the end-user. Another parameter is the reduction in throughput time of the system and minimization of the overhead of the proposed system. The developed system is then tested against some state-of-the-art baselines, and it is found that AGRICLOUD is performing better in a number of scenarios.

Table 1. Survey of various agriculture and classification-based approaches.

Reference	Year	Description	Advantages	Limitations
[19]	2015	System was formed to check the drought level	Higher accuracy Efficient High cohesion and low coupling	Limited/controlled environment Scalability issue
[20]	2017	The system performed several monitoring functions over the land	The intelligent robot performs the functionality in real-time	Low accuracy High cost
[21]	2015	The system can check the greenhouse effect over the covered land	Useful for low attitude level Uses 3D for analyzing data in real-time	Low sampling frequency rate High attitude level and wind speed constraints
[22]	2018	The system is developed to monitor the farm	Power-efficient system System is scalable Security incorporated	Limited functionality Data analytics required
[23]	2019	The system can check the problem of the crop and take action	System preempts disease illness Low power consumption	Less coverage area More computational time
[24]	2017	A combination of RF+SVM are used to classify tree species	Limited spectral bands	Datasets are few
[25]	2020	Invasive and expansive species classifications are performed using airborne hyperspectral data	Highest accuracy	Conditional constraints
[26]	2020	The smart drone is developed to check the land with few parameters	Limited functionality in a controlled environment Needs extensive testing	Intelligent features are missing Low accuracy
[27]	2021	A hybrid classification technique is developed for fruits classification	Best classification accuracy Useful for multiple technologies and platform	Multi-fruit classification can add more value to the system
[28]	2020	The Deep Convolution Neural Network approach is used to classify diseases in fruits	High accuracy, sensitivity, and precision values are found	Incremental computational time Complexity involved
[29]	2017	The system proposes classification method for the detection of plant diseases	Basic functionality is achieved	A number of parameters need to be added for better classification

3. Methodology

This section may be divided into subheadings. The data stored in the cloud computing environment are available to all agriculture advisors, researchers, and farmers. The advisors can extract agriculture data about farms applying the same pesticides and analyze the treatment suggested by other agriculturists and experts that helps improve decisions. Furthermore, it is observed that sudden pest attacks are always referred to as uncontrollable when expert treatments are obtained. However, to reach out towards experts when in critical condition is not easy, and as a result the whole crop becomes destroyed. The proposed approach facilitates the advisors/farmers working in rural areas to discuss problem cases with experts in real-time and to treat these advisors/farmers in a better way. The overall functionality of the analysis and monitoring system includes the steps below:

1. Farms are trained with a wireless sensor attached to specific points and a cellular device that is efficient enough to communicate through the Internet;
2. The module of the wireless sensor collects the farm's data and sends it to the mobile device via Bluetooth without user intervention;
3. A software of the client in the mobile device exhibits and transmits data to the analysis web service that is hosted by a cloud computing-based software. This interaction can occur with the mobile's data connectivity (e.g., mobile 4G/5G network), a home wireless gateway, or directly;
4. The analysis software carries out several computations over the collected data by taking reference from the current demographic data and the farm's historic data. Computations concern comparison, classification, and systematic diagnoses of farm

lines, which can be time-consuming especially when it is performed for longer periods and for a large number of farmers;

5. The software then extends the latest results to the historic record provided in private and secure cloud-based storage so that verified users can gain access to it anytime and from any place. Agriculturists then understand and translate the features extracted from the data, and a decision is made accordingly;
6. The results in the form of advice are disseminated to the farmer's mobile device;
7. The computing and monitoring processes are repeated daily/hourly, corresponding to the user's choice.

The proposed architecture is divided into three layers, such as the acquisition layer, the application layer, and the information exchange layer, as shown below in Figure 1.

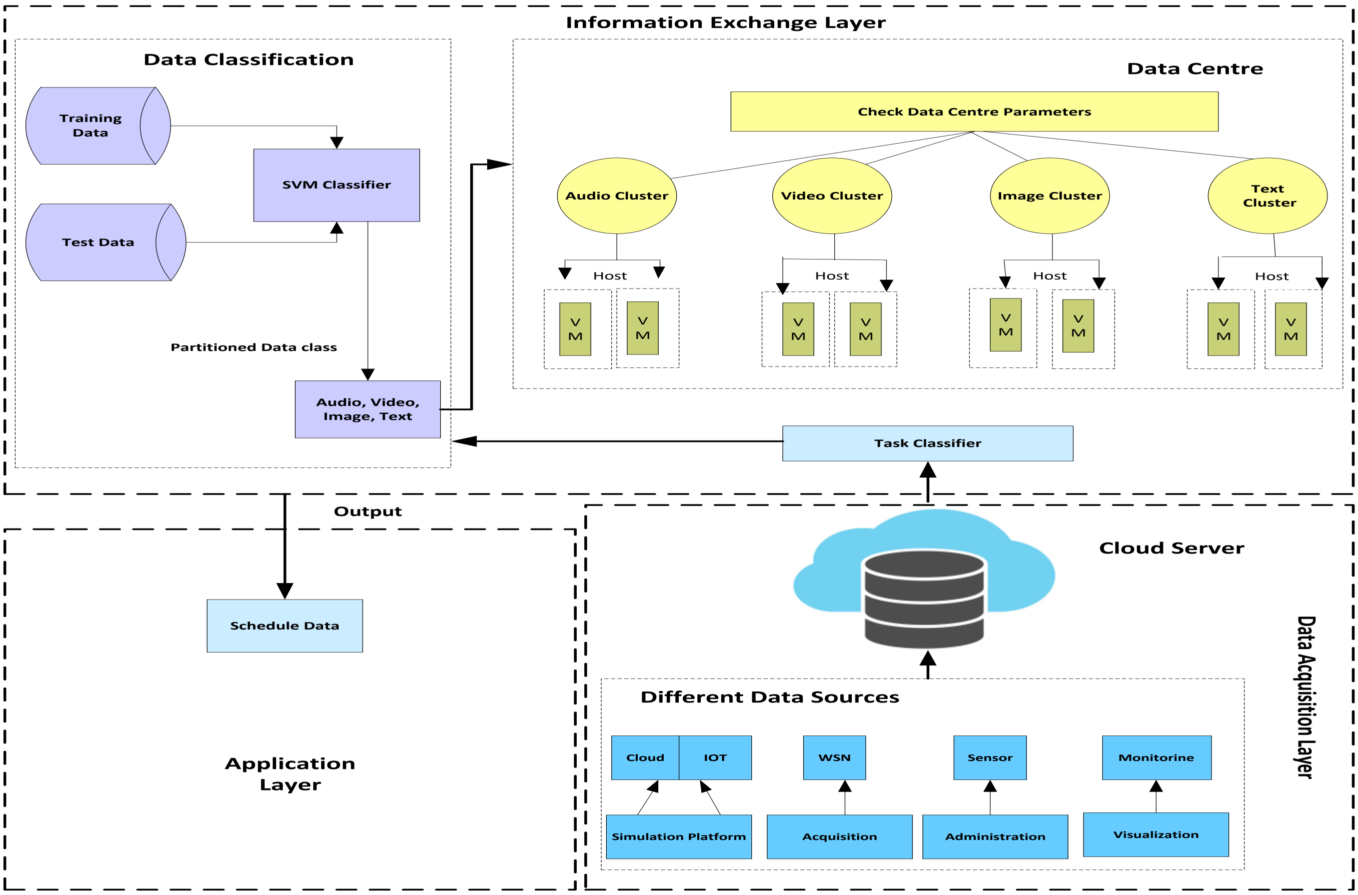

Figure 1. The flow of activities for the AGRICLOUD.

3.1. Acquisition Layer

The acquisition layer comprises various sensors used to collect data installed at different locations. The data received in data collecting devices are of numerous formats such as audio, video, images, maps, texts, etc. In order to perform timely computation and to reduced overhead, there is a need to implement the classification of data of those types.

For the development of cloud platform of the agriculture sector, it is important to implement steps such as the creation of an information model, understanding the ontologies, understanding the data, cataloging the data, building information model based on collection information, building service model, creating service model, understanding services, building service model, creating process model, understanding processes, converting services to processes, building process model, creating a governance model, define-design-implement policies, testing cloud architecture, creating a test plan with black and white box testing, selecting platform-deploy-processes-services-data, and assigning candidate data-services-processes.

3.2. Information Exchange Layer

The information exchange layer is responsible for sharing the data from the acquisition layer to the application layer through the datacenter and runtime information.

The role of a classifier is to establish which class, among a set of different classes, an unknown object belongs. A classifier has to find the best boundaries between the different classes. The optimal classifier, called SVM, minimizes the total cost. They are based on the separating hyperplane. If the data have high enough dimensionality, the different classes, which constitute different clusters, are linearly separable by hyperplanes.

An SVM model represents training data as points in space, which are then separated into categories by a wider gap. The classifier then finds the hyperplane that differentiates two classes and then it is categorized according to the gap they belong to. Furthermore, SVMs are effective in high dimensional spaces and in instances where the amount of dimensions is larger than the number of samples, and it is also memory efficient because it uses a subset of training data in the decision function which are called the support vectors. The kernel trick permits dealing with non-linear separating surfaces since the data, initially in the input space, are transposed in a higher dimensional space where a separating hyperplane is found. Figure 1 illustrates this principle.

We have divided the VMs into four types of sets, such as Audio VM, Video VM, Text VM, and Image VM based on input data. Each set of VM has different processing and storage resources in a cloud environment. More precisely, each machine (VM) is assigned a task based on task requirements. For example, video tasks require 1000 floating-point operations and 16 Giga Byte (GB) memory, audio tasks require 800 floating-point operations and 12 GB memory, image tasks require 800 floating-point operations, and 8 GB memory, textual tasks require 400 floating-point operations, and 4 GB memory. For video data classification, we extracted feature vectors of sequences of 40 frames extracted from four different video classes, where we have a 40 $\times$ 4096 matrix and where each row refers to features of one frame (one frame per row). Thus, we classified videos between these four different classes. We preprocessed a new video to limit its number of frames and then extracted features from this video to classify it. Assume that we have four video classes (ci, i = 1, . . . , 4). Each video has 40 (n = 1, . . . , 40) frames and from each frame we extracted 4096 features (1 $\times$ 4096). Since each frame has enough information to predict the video class (ci), we used 40 frames from each video as training/test samples, which creates an input matrix of (160 $\times$ 4096) dimensions, with 160 samples and each sample have 4096 features. Additionally, we have created an output vector (160 $\times$ 1) that contains the label of each class ci = i, where i = 1, . . . , 4. The feature sets include low-level signal properties, mel-frequency spectral coefficients, and two new sets based on perceptual models of hearing. For image classification, we have considered 256 $\times$ 256 pixels (total of 65,536 pixels). We used each pixel as a feature in the SVM classifier. For text classification, there are text documents of about 6 GB which are extracted in the form of unstructured text. We performed stemming and stopped word removal and extracted the words in the form of features. We then used these features for text data classification using SVM. The deep learning models require more time to train and their convergence time is greater, as compared to SVM; thus, so this is why we have adapted the traditional machine learning approach [30].

One of the problems of cloud task scheduling is assigning jobs to various VMs to accomplish load balancing in the shortest time. The benefit is that cloud resources are better used and users' requests are met more quickly. One of the major machine learning techniques is the Support Vector Machine (SVM), proved as a powerful model for achieving accurate classifications if carefully performed. In this case, we are proposing one-to-many types of SVM so that one type of data is taken as the positive class at a time and the other three classes are taken as the negative class [15]. For example, in the first case, audio data are classified from the rest of the data classes, then video data, image, map, and video data are classified separately. This preprocessing is performed within the cloud, which results in saving computations later in the information exchange phase. In one-to-many types of classification using SVM, multiple classes are separated from one another. A few of the examples include an email folder containing different categories such as work, family, and students. We have a classification problem in which four classes are used with assigning numbers, such as workshop x1 = 1, family matters x2 = 2, customers orientation x3 = 3, and students behavior x4 = 4. In binary classification, one class is taken as positive, and the rest of the classes are taken as negative. Here, the idea of the one vs. all classification is taken to obtain the best accuracy results since this study involves four training sets with class 1 video data type as z1 = 1, class 2 text data type as z2 = 2, class 3 audio data type as z3 = 3, and class 4 image data type as z4 = 4. The proposed method is to take one class at a time and to turn it into four separate binary classifications. In this case, class 1 is taken as a positive class with value 1 and the three other classes as a negative class with value 0. When we train the datasets, this will provide us with a decision boundary of one class with the rest of the classes. In simple words, in the first instance, video datasets are taken as positive classes and all others as negative classes. In the second instance, text datasets are taken as a positive class, and the rest are negative classes. In the third instance, audio datasets are taken as positive classes, whereas others are negative classes. Finally, in the last instance, images belong to the positive class, whereas all other datasets are taken as negative classes. To make the communication secure, the AES lightweight version of the protocol is suggested so that information at the time remains secure and cannot be changed by the attacker in any type of security active attack. This helps in achieving confidentiality, integrity, and authentication of the exchanged information even within the layers until the end-users.

3.3. Application Layer

The application layer is responsible for displaying the results to the end-users in the form of GUI. These results consist of schedule data.

Algorithm 1 describes the steps involved in the data partition and data scheduling process.

Algorithm 1. AGRIClOUD

Input: video, text, audio, digital maps from different cloud sources, number of virtual machines (VM)
Output: Data class, Scheduled data

for data classification **do**
  for each $P(u,v)$ **do**
    Evaluate ← Kernel SVM
    for each Classification accuracy ≠ M **do**
      Evaluate data accuracy
      if Number of iterations ≠ N **then**
        perform data categorization (data partition)
return data class
for load balancing **do**
  schedule data ← Assign data to each group (cluster) w.r.t virtual machine
end for
return schedule data

We used one of the simplest kernels. The space X corresponds to R^N, and we assume patterns x in which most information can be represented by the dth order products of entries $[x]j$ of x. These products are called monomials. Here, the vector of monomials is used to represent the feature space H, and the mapping itself represents Φ. One can easily show that the dot product in the feature space is simply the square of the dot product in the input space. The hyperfunction is as follows:

$$K(u_i, u_j) = \varphi(u_i)\varphi(v_j) \quad (1)$$

$$f(u_j) = \sum_{i=1}^{N} \propto_i u_i K(u_i u_j | u_i) + c. \quad (2)$$

where u_i. is used for support vector, $\propto_i$ is represented as Lagrange multiplier, and u_j is known as the label of membership class (+1, −1) where n = 1,2,3, ... N.

4. Results

To check the functionality of the proposed system, the CloudSim4.0 simulator developed by University of Melbourne, Australia is used at the cloud server. The CloudSim4.0 simulator [31] is devised to perform the working of the systems in the simulated form under different scenarios. We have set the following parameters for our proposed system.

Overall, 100,000 files are collected at run time in various formats such as audio, video, image, text, and maps. These files are of varying sizes from 0.1 MB to 1 GB. The configuration of data centers, their size, number of hosts, Random Access Memory (RAM) size, task size on disk, storage capacity, bandwidth, and power are shown in Table 2. The simulation is performed on a Core i9 Intel Quad-core system (by Intel Corporation, USA) with 16 GB RAM and 4 TB HDD. Data files numbering 100,000 after preprocessing using SVM classification have been divided into an equal number of audio, video, image, and texts formats shown in Table 3. This helps in further processing the data with low computation cost and high accuracy as mentioned in Table 4. Cloud computing is making significant contributions in extracting useful information when combined with data distribution techniques. There could be another choice, such as Spark in data distribution, which makes it easy for huge volumes and a variety of data with the help of load balancing approaches. However, we have designed a model with minimum hardware cost and information processing cost for a large amount of data. Although Apache Spark (developed by UC Berkeley, USA) may provide better running time, our basic objective was to design an optimized model using machine learning algorithms.

Table 2. Parameters used for AGRICLOUD in the CloudSim4.0.

Datacenter	Parameters	Values	Parameters	Values
	Datacenter's Size	4–16	RAM size	8–32 GB
	Hosts in the Datacenters	8–32	Storage capacity	1–4 TB
Tasks	RAM size for Host (each)	4–16 GB	Data BW(Band Width)	5–15 GB/s
	Total tasks in a number	100,000	PE Power	108 MIPS
	Size of each task	0.1 MB–1 GB	Tasks size	50 GB
VM	VM ID	(4–400)	VMM	Xen

Table 3. Training and Test sets details.

Samples	Size in Quantity	Samples	Size in Quantity
Training Audio (Datasets)	17,500	Testing Audio (Datasets)	7500
Training Video (Datasets)	17,500	Testing Video (Datasets)	7500
Training Images (Datasets)	17,500	Testing Images (Datasets)	7500
Training Text (Datasets)	17,500	Testing Text (Datasets)	7500
Total Data files		100,000	

Table 4. Comparative analysis of AGRICLOUD with Classification Techniques.

Metric	AGRICLOUD	SVM-GA	MSVM	MULTICLASS	ACOSVM
Accuracy	0.874	0.872	0.789	0.821	0.781
Precision	0.873	0.863	0.798	0.822	0.791
Recall	0.871	0.874	0.722	0.862	0.839
F-Measure	0.877	0.873	0.774	0.883	0.861
G-Mean	0.870	0.853	0.754	0.852	0.799
AUC	0.871	0.851	0.762	0.851	0.819

The confusion matrix in terms of accuracy, precision, recall, F-measure, G-mean, and (Area Under Curve) AUC is calculated for the baselines classifiers such as SVM-GA [32], MSVM [33], MULTICLASS (Multiple Class) [34], and ACOSVM (Ant Colony Optimization using Support Vector Machine) [35]. All these baselines have proved their performance against the number of algorithms in the literature. However, after performing preprocessing using SVM one-to-many types, our AGRICLOUD has performed extremely well and has shown better classification performance. This will help in the further minimization of execution time later in the completion phase that results in benefiting the end-user such as a farmer, researcher, adviser, or expert.

5. Discussion

The developed algorithm AGRICLOUD is compared with CLAYMIST (cloud enabled CMM index for smart agriculture monitoring system) [36] and MISSENARD [37] baselines in terms of execution time, throughput time, and overhead time. In the first case, execution time is demonstrated for all algorithms in Figure 2. The number of tasks in the form of audio, video, images, texts, and maps is executed gradually over 5, 10, 20, 50, 100, 500, and 1000 VMs. It can be observed that, from the very start, AGRICLOUD remains very stable in processing the tasks. However, at 100 and more VMs, more execution time is required, showing less performance, but its performance is still better than CLAYMIST and MISENARD. These two algorithms performed somehow equal to 500 VMs but gradually started to require more execution time, rendering them the worse choice. This shows that the proposed algorithm has better classification performance and scalability than others.

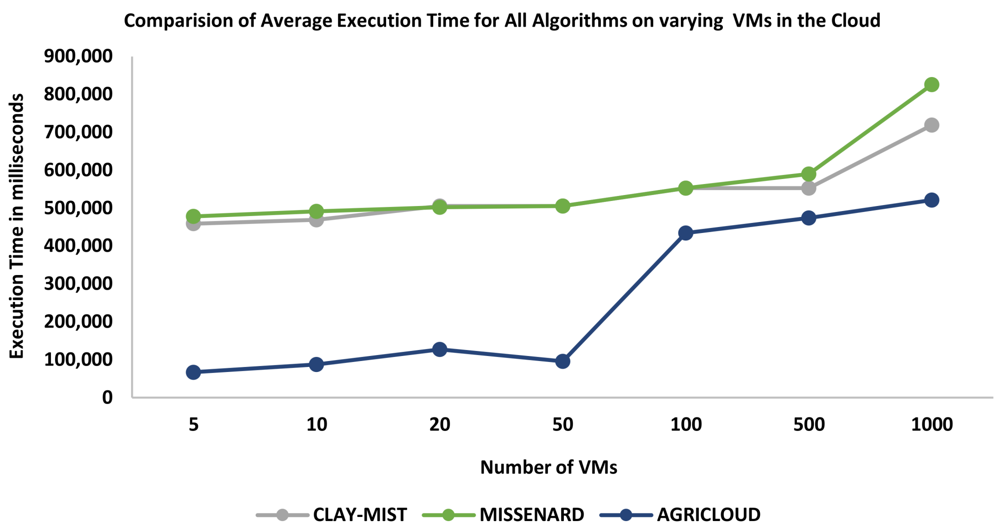

Figure 2. Execution time of AGRICLOUD with baselines.

Figure 3 shows throughput time taken by the baselines over 5, 10, 20, 50, 100, 500, and 1000 VMs. It is clear that all algorithms start better at the beginning and gradually increase their time when several tasks and VMs increases. At the end of the last simulation over 1000 VMs, AGRICLOUD performed comparatively better than MISSENARD and CLAYMIST. This shows that the proposed algorithm can make responses fast in the execution of the tasks.

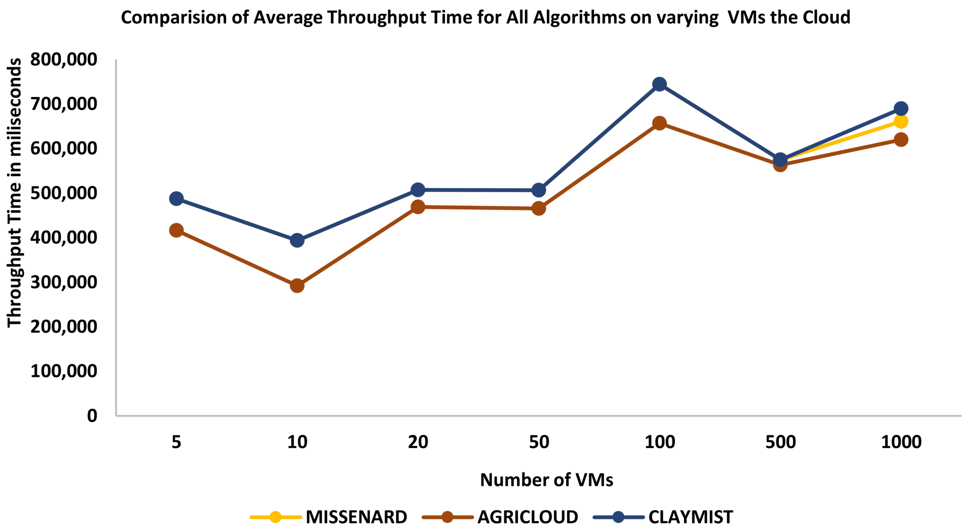

Figure 3. Throughput time of AGRICLOUD with baselines.

Figure 4 shows the overhead time taken by all baselines over a set of VMs. The lower the overhead, the faster the response time and the higher the execution time will be. All algorithms start well at the beginning with low tasks and a smaller number of VMs. The algorithms abruptly started to take more overhead time when it reaches 100. However, due to the earlier performance of the INDIGSOL for better execution and throughput time, the overhead time remains slow, showing better performance of the proposed algorithm. Overall, this would help in processing the data in the form of audio, video, text, images, and maps much faster for the end-users. The end-user would be able to receive critical information such as weather, temperature, climate change, pesticide information, crop production, water level, soil moisture, and several other factors timely and precisely and with accuracy. This information is stored in the cloud in the form of historical data that would be updated continuously with every packet of information. Furthermore, the information is beneficial to the advisors, experts, and especially to the farmers at their doorstep without any additional cost. Table 5 shows the statistical data of AGRICLOUD with different baselines.

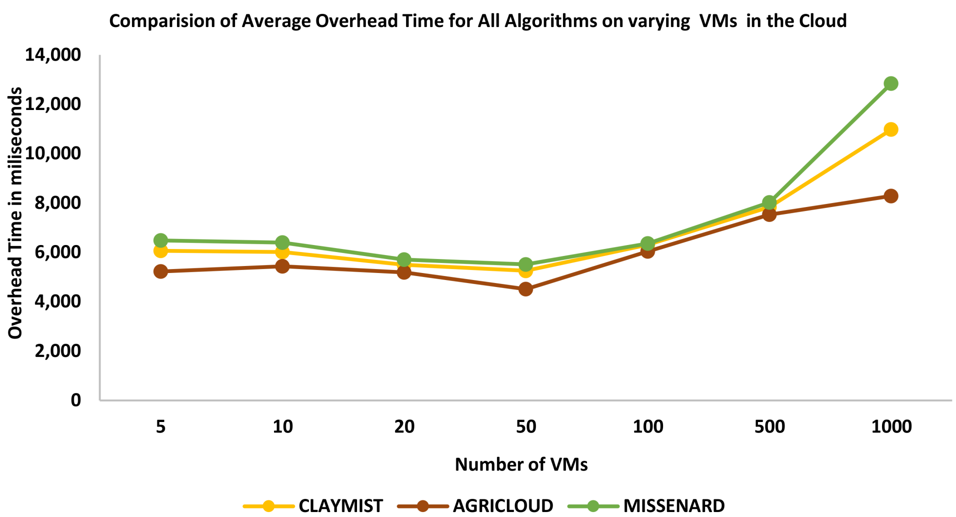

Figure 4. Overhead time of AGRICLOUD with baselines.

Table 5. Statistical analysis of AGRICLOUD with other baselines.

Statistical Parameters			
Execution Time	MISSENARD	CLAYMIST	AGRICLOUD
(Standard Deviation) SD	9.466202262	7.624263468	3.162963415
Mean	61.46153846	59.55128205	33.9631641
p-value	1.3279E-05	0.000331995	
t-value	1.61703E-09	2.9834E-10	
Overhead Time	CLAYMIST	MISSNARD	AGRICLOUD
(Standard Deviation) SD	5394.790954	4968.464232	1466.939567
Mean	4414.383949	4430.255744	2917.987817
p-value	0.001200862	0.001673218	
t-value	0.033681748	0.047162288	
Throughput Time	MISSNARD	CLAYMIST	AGRICLOUD
(Standard Deviation) SD	785437.3736	566543.8203	259145.3555
Mean	633933.8473	525730.1763	314728.5988
p-value	1.58317E-06	6.00133E-08	
t-value	0.019134393	0.007580034	

Student's *t*-test is used to find out the significant difference between AGRICLOUD and two baselines. This test finds out the difference between two independent groups that help to determine whether the difference is actual or due to chance. It is shown that the *p*-value for execution time, overhead time, and throughput time for all baselines is greater than 0.01 significance level. Similarly, the t-value for all baselines is greater than 0.01, showing that a significant difference exists. Therefore, statistically, our proposed algorithm is better than other baselines as well.

6. Conclusions and Future Work

Pakistan, being an agricultural country with 70% land, needs to pay much focus on the adoption of new technologies concerning global practices. The motivation assists our farmers in becoming advanced at the supervisory level so that they can manage things appropriately with an understanding of the farms. There is a need to build the capacity of the farmers so that they interact with one another in addition to experts. Currently, no such mature work exists both in practice and at the infrastructure level. The cost of management is the most important consideration here. As a result, the scope has been limited to computer resources, such as storage, processing, and data transport.

A farm health monitoring system now uses a mobile phone's processing capability, integration with farm sensors, and internet access to alert farmers when the crop requires attention or even perform an automatic intervention, such as triggering the automatic release of drugs into the farm when necessary. We designed a cloud computing-based real-time agricultural monitoring and analysis system capable of assisting farmers in better managing their farms by decreasing or eliminating on-site consultations.

An AGRICLOUD is proposed to solve most of the aforementioned issues currently present in agriculture. The main goal of this research is to propose an innovative solution that performs file formatting in the cloud so that pre-processing of a huge volume and variety of files will be performed initially. This pre-processing will save time when resources are assigned to the virtual machines resulting in much-reduced execution time and energy consumption. This approach makes the solution viable, robust, and scalable as it has been tested with gradually more and more resources. It is important to mention that file formatting in the cloud for agricultural data is an innovative step in smart farming that contributes to the body of science. It is observed that in some current baselines, the proposed system performed well for a few parameters such as execution time, throughput time, and overhead time while reducing the energy.

The authors are residents of Pakistan where most of the land is agricultural and belongs to rural areas. Here, people have low income and require innovative technologies to improve their farming skills so that per capita income may increase. The Government of Pakistan has initiated few agricultural reforms in which smart farming is one of them. Initially, this study is a pilot test study and is implemented with only Pakistan climate factors considered due to the huge cost involved. However, this study will also be extended to other countries in the future.

In the future, we will work on critical factors of cloud-based agriculture such as priority-based farming, agriculture 5.0, precision agriculture, telematics, and data analytics by using artificial intelligence-based techniques. This would help in making optimal decisions by keeping in view the availability of a certain number of resources.

Author Contributions: Conceptualization, M.J., A.S., and M.U.H.; methodology, A.A. and K.R.; software, M.S.A.R.; validation, M.A., M.J., and A.S.; formal analysis, M.J. and A.S.; investigation, M.U.H.; resources, K.R.; data curation, M.S.A.R. and M.A.; writing—original draft preparation, M.J. and A.S.; writing—review and editing, M.U.H. and A.A.; visualization, M.S.A.R.; supervision, K.R.; project administration, M.U.H. and M.A.; funding acquisition, A.S. All authors have read and agreed to the published version of the manuscript.

Funding: The authors would like to acknowledge the support of the Deputy for Research and Innovation-Ministry of Education, Kingdom of Saudi Arabia, for this research through a grant (NU/IFC/INT/01/008) under the institutional Funding Committee at Najran University, Kingdom of Saudi Arabia.

Data Availability Statement: Data are within the article.

Acknowledgments: The authors would like to thank all advisors and teachers for their support.

Conflicts of Interest: The authors declare that there are no conflicts of interest.

References

1. Husain, I. Pakistan's economy and regional challenges. *Int. Stud.* **2018**, *55*, 253–270. [CrossRef]
2. Gaitan, A.; Gaitan, V.G.; Ungurean, I. A survey on the internet of things software architecture. *Int. J. Adv. Comput. Sci. Appl.* **2015**, *6*, 140–143.
3. Mekala, M.S.; Viswanathan, P. A novel technology for smart agriculture based on iot with cloud computing. In Proceedings of the International Conference on IoT in Social, Mobile, Analytics and Cloud (I-SMAC), Palladam, India, 10–11 February 2017; pp. 75–82.
4. Junaid, M.; Sohail, A.; Rais, R.N.B.; Ahmed, A.; Khalid, O.; Khan, I.A.; Hussain, S.S.; Ejaz, N. Modeling an Optimized Approach for Load Balancing in Cloud. *IEEE Access* **2020**, *8*, 173208–173226. [CrossRef]
5. Markovic, D.; Koprivica, R.; Pesovic, U.; Ranic, S. Application of IoT in monitoring and controlling agricultural production. *Acta Agric. Serbica* **2015**, *12*, 145–153. [CrossRef]
6. Wu, S.; Austin, A.; Ivoghlian, A.; Bisht, A.; Wang, K. Long range wide area network for agricultural wireless underground sensor networks. *J. Ambient Intell. Humaniz. Comput.* **2020**. [CrossRef]
7. King, I. Technology: The future of agriculture. *Nature* **2017**, *544*, S21–S23. [CrossRef] [PubMed]
8. Buyya, R.; Yeo, C.S.; Venugopal, S. Market-oriented cloud computing: Vision, hype, and reality for delivering IT services as computing utilities. In Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications, Dalian, China, 25–27 September 2008; pp. 5–13.
9. Nurmi, D.; Wolski, R.; Grzegorczyk, C.; Obertelli, G.; Soman, S.; Youseff, L.; Zagorodnov, D. The eucalyptus open-source cloud-computing system. In Proceedings of the 9th IEEE/ACM International Symposium on Cluster Computing and the Grid, Shanghai, China, 18–21 May 2009; pp. 124–131.
10. Foster, I.; Kesselman, C. *The Grid: Blueprint for a New Computing Infrastructure*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1998.
11. Adetunji, K.E.; Joseph, M.K. Development of a cloud-based monitoring system using 4Duino: Applications in agriculture. In Proceedings of the International Conference on Advances in Big Data, Computing and Data Communication Systems, Durban, South Africa, 6–7 August 2018; pp. 4849–4854.
12. Lu, H.; Ji, C.; Yan, J. An IoT and mobile cloud based architecture for smart planting. In Proceedings of the 3rd International Conference on Machinery, Materials and Information Technology Applications (ICMMITA), Qingdao, China, 28–29 November 2015; pp. 1001–1005.
13. Iosup, A.; Ostermann, S.; Yigitbasi, M.N.; Prodan, R.; Fahringer, T.; Epema, D. Performance analysis of cloud computing services for many-tasks scientific computing. *IEEE Trans. Parallel Distrib. Syst.* **2011**, *22*, 931–945. [CrossRef]
14. Mergen, M.F.; Uhlig, V.; Krieger, O.; Xenidis, J. Virtualization for high-performance computing. *ACM SIGOPS Oper. Syst. Rev.* **2006**, *40*. [CrossRef]

15. Narasimhan, P.; Nichols, R. State of cloud applications and platforms: The cloud adopters view. *Computers* **2011**, *44*, 24–28. [CrossRef]
16. Thornton, P.; Dinesh, D.; Cramer, L.; Loboguerrero, A.M.; Campbell, B. Agriculture in a changing climate: Keeping our cool in the face of the hothouse. *Outlook Agric.* **2018**, *47*, 283–290. [CrossRef]
17. Banerjee, P.; Friedrich, R.; Bash, C.; Goldsack, P.; Huberman, B.; Manley, J.; Patel, C.; Ranganathan, P.; Veitch, A. Everything as a service: Powering the new information economy. *Computer* **2011**, *44*, 36–43. [CrossRef]
18. Lin, C.; Lu, S.; Lai, Z.; Chebotko, A.; Fei, X.; Hua, J.; Fotouhi, F. Service-oriented architecture for view: A visual scientific workflow management system. In Proceedings of the 2008 IEEE International Conference on Services Computing, Honolulu, HI, USA, 7–11 July 2008; pp. 335–342.
19. Luan, Q.; Fang, X.; Ye, C.; Liu, Y. An integrated service system for agricultural drought monitoring and forecasting and irrigation amount forecasting. In Proceedings of the 23rd International Conference on Geoinformatics, Wuhan, China, 19–21 June 2015; pp. 1–7.
20. Krishna, K.L.; Silver, O.; Malende, W.F.; Anuradha, K. Internet of things application for implementation of smart agriculture system. In Proceedings of the International Conference on I-SMAC, Palladam, India, 10–11 February 2017; pp. 54–59.
21. Malaver Rojas, J.; Gonzalez, F.; Motta, N.; Villa, T.F.; Etse, V.; Puig Garcia, E. Design and flight testing of an integrated solar powered UAV and WSN for greenhouse gas monitoring emissions in agricultural farms. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Hamburg, Germany, 28 September–3 October 2015; Volume 1, pp. 1–6.
22. Davcev, N.; Mitreski, K.; Trajkovic, S.; Nikolovski, V.; Koteli, N. IoT agriculture system based on LoRaWAN. In Proceedings of the 14th IEEE International Workshop on Factory Communication Systems (WFCS) 2018, Imperia, Italy, 13–15 June 2018; pp. 1–4.
23. Santos, D.; Pessin, G.; Costa, D.; Righi, D.R. Agri prediction: A proactive internet of things model to anticipate problems and improve production in agricultural crops. *Comput. Electron. Agric.* **2019**, *161*, 202–213. [CrossRef]
24. Raczko, S.; Zagajewski, B. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *Eur. J. Remote Sens.* **2017**, *50*, 144–154. [CrossRef]
25. Tomala, S.; Raczko, E.; Zagajewski, B. Comparison of support vector machine and random forest algorithms for invasive and expansive species classification using airborne hyperspectral data. *Remote Sens.* **2020**, *12*, 516. [CrossRef]
26. Namani, S.; Gonen, B. Smart agriculture based on IoT and cloud computing. In Proceedings of the 3rd International Conference on Information and Computer Technologies (ICICT), San Joe, CA, USA, 9–12 March 2020; pp. 553–556.
27. Nasir, I.M.; Bibi, A.; Shah, J.H.; Khan, M.A.; Sharif, M.; Iqbal, K.; Nam, Y.; Kadry, S. Deep learning-based classification of fruit diseases: An application for precision agriculture. *Comput. Mater. Contin.* **2021**, *66*, 1949–1962. [CrossRef]
28. Akram, T.; Sharif, M.; Saba, T. Fruits diseases classification: Exploiting a hierarchical framework for deep features fusion and selection. *Multimed. Tools Appl.* **2020**. [CrossRef]
29. Elangovan, T.; Nalini, S. Plant disease classification using image segmentation and SVM techniques. *Int. J. Comput. Intell. Syst.* **2017**, *13*, 1821–1828.
30. Senekane, M. Differentially Private Image Classification Using Support Vector Machine and Differential Privacy. *Mach. Learn. Knowl. Extr.* **2019**, *1*, 483–491. [CrossRef]
31. Svorobej, S.; Takako Endo, P.; Bendechache, M.; Filelis-Papadopoulos, C.; Giannoutakis, K.M.; Gravvanis, G.A.; Tzovaras, D.; Byrne, J.; Lynn, T. Simulating fog and edge computing scenarios: An overview and research challenges. *Future Internet* **2019**, *11*, 55. [CrossRef]
32. Girsang, A.S.; Manalu, A.S.; Huang, K. Feature selection for musical genre classification using a genetic algorithm. *Adv. Sci. Technol. Eng. Syst. J.* **2019**, *4*, 162–169. [CrossRef]
33. Dubey, S.R.; Jalal, A.S. Apple disease classification using color, texture and shape features from images. *Signal Image Video Process.* **2016**, *10*, 819–826. [CrossRef]
34. Kanu, P.; Vala, J.; Jaymit, P. Comparison of various classification algorithms on iris datasets using WEKA. *Int. J. Adv. Eng. Res.* **2014**, *1*, 1–7.
35. Junaid, M.; Sohail, A.; Ahmed, A.; Baz, A.; Khan, I.A.; Alhakami, H. A hybrid model for load balancing in cloud using file type formatting. *IEEE Access* **2020**, *8*, 118135–118155. [CrossRef]
36. Mekala, M.S.; Viswanathan, P. CLAYMIST: IoT-cloud enabled CMM index for smart agriculture monitoring system. *Measurement* **2019**, *134*, 236–244. [CrossRef]
37. Ray, P.P. Internet of things cloud enabled MISSENARD index measurement for indoor occupants. *Measurement* **2016**, *92*, 157–165. [CrossRef]