Federated Learning: Crop Classification in a Smart Farm Decentralised Network.

Godwin Idoje[a,1,*], Tasos Dagiuklas, Iqbal Muddesar

[a]*Computer Science and Informatics division, London South Bank University, United Kingdom*

Abstract

In this paper, the application of federated learning to smart farming has been investigated. The Federated averaging model has been used to carry out crop classification using climatic parameters as independent variables and crop types as labels. The decentralised machine learning models have been used to predict chickpea crops. Through experimentation, it has been observed the model converges when learning rates of 0.001 and 0.01 are considered using the Stochastic gradient descent (SGD) and the Adam optimizers. The model using the Adam optimizer converged faster than the SGD optimizer, this was achieved after 100 epochs. Analysis from the farm dataset has shown that the decentralised models achieve faster convergence and higher accuracy than the centralised network models.

Keywords: Federated Learning, classifier chain Gaussian (CCGNB), Binary Relevance Gaussian (BRGNB), Label powerset Gaussian Naïve Bayes (LPGNB)

1. Introduction

The Federated Learning (FL) approach has been adopted in this research to ascertain how the predicted crop types are close to the original crop types within the provided dataset. An FL network is a decentralised network where the local edge nodes send their updated weights to the server and the server aggregates all these updated weights and sends the combined model back to the edge nodes for further training, this process continues until convergence is achieved. The models are hyper-tuned to investigate the convergence of the decentralised models during the optimization of the Federated Averaging model of the smart farming dataset. The FL algorithm is used to aggregate the edge node models within the decentralised network. The FL server sends its base model to the edge nodes and these edge nodes use the base model for training its local datasets and send their updated weights to the server, the server aggregates all the various updated weights from each edge node and forms a new global model, the updated global model is sent back to the edge nodes for the local model training, this process continues until convergence is obtained. As discussed in [1], the use of satellite images has helped to analyze soil and crops in farmlands, to determine the condition of the crops or soil. This has helped to resolve many challenges with soil and crops

[*]Author correspondence

[1]Email: idojeg2@lsbu.ac.uk

[2]www.suitelab.org

[3]Computer science, London South Bank University,103 Borough Road, London, SE1 0AA, UK

using data obtained via satellite, these solutions have enhanced farming through the forecast of the crops harvesting time to make decisions to combat poor harvest from the farms. According to [2], six domain models have been used for designing smart farms to interconnect between systems, the domain models have enhanced the joint ecosystem of sharing data between the industry players. It can be inferred that smart farms use Information and Communication Technology (ICT) and Internet of Things (IoT) devices which are interconnected via the Internet. Many data can be exchanged within these farms and one of the challenges experienced in smart farming is data privacy. It will be interesting to investigate federated learning applications for these smart farms, where the data owner will not share the data with the data scientist. Therefore, the data scientist will be able to evaluate these smart farm datasets without access to the farm data, this research explores smart farming use case within the federated learning platform.

The research of [3] discussed the automation of a smart farm where the irrigation system is controlled via a mobile app, enabling farmers to monitor the data captured by the IoT device around the plant. The limitation of their work is that their system has not performed any analysis of the data collected. The research conducted by [4,5,6] discussed predictions of pest infestation from the dataset captured from smart farms. The plants experience high moisture during the day and low moisture at night, prompting the researchers to calibrate the automated device to supply more water to the plants during the day. The limitation of their work is that they did not provide any analysis of the data captured. It was observed from [1, 2, 3, 4, 5] that many of the edge devices performed little or no analysis of the captured data. The synergies of the results, predictions, or analysis have not been achieved due to different platforms, and operating systems used in their research. This paper uses a multi-labelled dataset, the climatic parameters are the independent variables while the crop types are the labels and the federated learning platform has been used to predict the crop types from the climatic parameters for a smart farm network.

Our Contributions: This paper proposes the use of hyper-tuned federated averaging models that can provide privacy for the smart farming multi-labelled dataset during evaluation. This is due to the fact the dataset is not shared with the server but is trained locally at the edge nodes. The FL models outperform the centralised network machine learning Gaussian Naïve Bayes models by producing optimal convergence, accuracy, and harmonic means. Therefore, the model can predict the crop type from the dataset which contain climatic parameters as independent variables and crops as the labels. The hyper-tuned Federated averaging algorithm has been able to make crop predictions with a high accuracy value from the given multi-labelled dataset without access to the raw data within a decentralised network. The climatic features of the data have been temperature, humidity, the potential of hydrogen(pH), and rainfall while the following crops have been the dependent variables rice, maize, and chickpea. A Testbed using PySyft, Pytorch and Syft libraries has been used for the emulation. Section 2 discusses the existing publications on federated learning. Section 3 narrates the methodology adopted using federated Learning to predict crop types from the climatic datasets. Section 4 discusses the results obtained from the federated learning model and the Gaussian Naïve Bayes classifier models for the multi-labelled dataset. Section 5 discusses the conclusion drawn from the results in Section 4.

2. Related work

2.1. *Federated Learning*

The soil-less smart farming methodology [7] has been adopted to cultivate crops, this approach has enabled farmers to produce high-yield crops, reduce water usage within the smart farm, and low parasite infestation. However, the smart farming technic affords farmers the opportunity to monitor their crops using IoT sensors. Federated learning (FL) is a machine learning technic for analysing datasets without accessibility to the raw data. According to [8], FL has been used in the medical industry for covid-19 disease detection using chest computed tomography images. Their result indicates that using the Federated averaging model, the communication cost of their network has been reduced. The authors in [9] propose a modified Federated learning model where the edge nodes are randomly distributed into groups, and a group is given a different transmission time slot, this technic has been able to reduce the Byzantine attacks within their Federated learning test bed. Edge devices are agnostic in their capacities and resources, [10] paper proposes a federated learning framework that accepts the ad-hoc nature of the edge devices and analyses the models without compromising the privacy, and security of the data while achieving convergence. Research papers [11, 12] propose the decoupling of the federated Learning architecture while distributing the edge nodes task in an intelligent pattern. This architecture leads to low computational resource usage. Edge devices have been used to capture data which are processed or transmitted to the cloud or server for analytics to ascertain decisions in various sectors. Data privacy is ensured during the evaluation.

The authors in [13] have discussed that recent research work in federated learning has discussed extensively supervised learning and they have suggested that researchers should consider investigating unsupervised machine learning within a federated learning platform. As discussed in [14], to preserve the privacy of the data trained in a machine learning system, a shift from the classical machine learning algorithm to a decentralised machine learning platform is important, where the data are not sent to the server or cloud for training, this equally reduce the latency since the bandwidth consumption within the network is reduced equally. It can be inferred from [12,13,14] that their work has not been applied to an unsupervised learning algorithm which is a limitation. As cited in [15], federated learning has been used to establish cross-domain, cross-data, and cross-enterprise platforms. The limitation of their research is that their work did not mention if they used either homogeneous or heterogeneous datasets. Homogeneous edge nodes all have the same attributes such as the memory, processor, and bandwidth capacities as opposed to heterogeneous edge nodes. In this paper's research, the edge nodes are homogeneous because all the edge nodes have the same memory, processor and transmitting power. The authors in [16,17] have discussed that there exists a server and edge nodes correlation and cross-domain, cross-data transaction between edge nodes and server nodes in a Federated learning network. It has been discussed that sending only the updated weights within the FL network minimised the latency within the network. It can be inferred that communication cost has been reduced by two ranks in an FL network from their research. Their research considered low bandwidth consumption edge nodes during the rounds but their model has not been tested in a high bandwidth scenario.

In [17], it is considered that a modest assets scenario in an FL network, Federated Distillation (FD), is an algorithm that reduces communication overhead better than the Federated averaging algorithm and Hybrid federated distillation (HFD) algorithm. This helps to enhance the performance gap between FL and FD by controlling the average probability vector and average input from the dependent variable during the offline phase. It was reported in their paper that FD and HFD yield better results compared with federated averaging when the number of uplinks and downlinks channels is very small. However, their research did not address the use of their model for a wired non-fading channel link and no information was provided on the frequency of the wireless edge nodes which was used for the experiment. The work of [16] inspires our architecture where a server has been set up for experimentation using homogeneous Edge nodes with the same attributes such as memory capacity, and processor. In [18], it is observed that using the distance of convex functions enables researchers to pick more nodes compared to other technological technics when the accumulation of Multi-access Edge Computing (MEC) devices allow applications to be run close to the service user for a rather demanding mean square error request which was achieved through increment of antennas at a base station in their experiment. The MEC allow cloud computing features and information technology profiles at the edge of any network. It can be deduced from their paper that aggregation of more MEC edge nodes in their experiment enhanced the performance of their model, some limitations were observed in their research such as, it did not investigate the effect of channel uncertainty in the model accumulation, more so their research did not address the computational complexity of the algorithm used. This paper applies the FL approach to smart farming, the Federated learning technique is a subset of machine learning that can be regarded as a contribution. The authors in [19] have used a greedy algorithm, a two-magnitude image analytical solution, where the edge nodes are vehicular. It can be deduced that the greedy algorithm helps to achieve model accuracy and aggregation efficiency for a federated learning vehicular network. Their work inspires the performance of FL models for smart farming. In this research, as shown in Figure 1, which depicts our architecture, the mobile edge computers receive the data from the IoT devices, the MEC perform the local training of the data and only sends their updated weights to the server, upon completion of the aggregation of all the received local weights, the server sends its new updated global model to each MEC and they also use this new received updated global model to perform the next training, this process continues until convergence is achieved.

The authors in [20] discussed, their modified C-fraction Federated Stochastic gradient descent algorithm which considers the ratio of the online participants to the total number of participants within the federated network, their modified algorithm has been able to give between 99.65% to 99.85% accuracy from the training using different values of the c-fraction during experimentation, despite the impressive results from their experimentation, it can be observed that the same learning rate has been considered for the 4 different C-fraction, it would have been interesting to get the results for each C-fraction using different learning rates. Many different learning rates have been considered for this research unlike the research of [20] to determine the effect of the different learning rates on our accuracy values using different optimizers such as Stochastic Gradient Descent (SGD) and Adam optimizer. According to the authors in [18], the Adam activation function has been used in a Federated averaging algorithm for a crowd-sourcing speech data to study an asset-limited wake word detector instead of using the normal global averaging for its training, their work achieved a 95% recall per 5 false alarm per hour (FAH) for 100 communication cycle when the crowd-sourced dataset communication cost per participant was 8 Megabyte (MB). Using the Adam optimisation, the network can converge faster, the limitation of their work is that a memory-efficient end-to-end model was not used in their research. [20] discussed that SGD converges faster but the step sizes decay fast which affects its efficiency during training, however, [21] stated that the Adam optimizer is a robust optimizer that combines two other optimizers namely Adagrad and RMSProp, and uses less memory for training and converges faster than SGD. This paper has considered both the SGD and Adam optimizer in our research for analysis and our results depict the performance of the model using smart farming variables within a Federated Learning network, the results indicate that the Adam optimizer had a higher accuracy compared with the SGD optimizer while using climatic variables for crop type prediction. It is obvious from [14-21] that federated learning has been implemented in various networks with edge nodes which have reduced edge node queuing, bottleneck traffic, and latency of traffic due to the application of different technic of algorithm schemes to make the communication cost low and the network more efficient. Related works have shown that several technics have been adopted by researchers to reduce the latency and network traffic challenges within a particular network, this research explores options for hyper-tuning the parameters to achieve optimal convergence within the federated learning network while predicting the crop type.

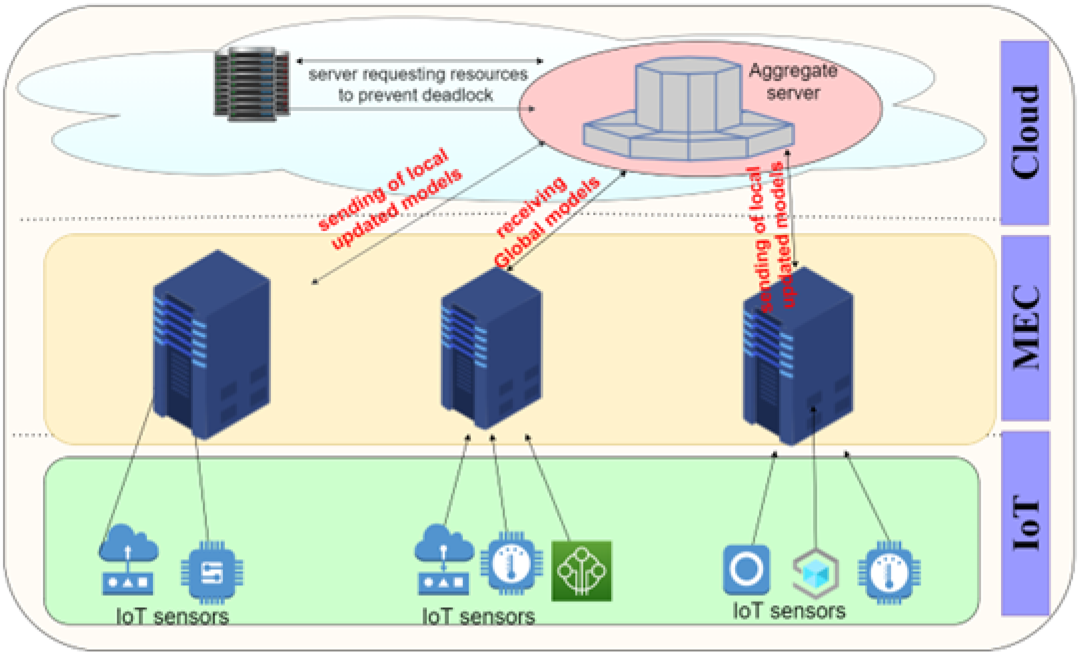

Figure 1: Federated Learning Architecture.

Figure 2 shows the Federated Learning network flow sequence from the sensors which capture data and send these data to the edge devices. Unlike classical machine learning where the data are sent to the cloud for training, Federated learning adopts a different approach, the server sends its initial global models to the edge devices. Since training takes place at the edge nodes where the data is domiciled, the edge devices use the initial global model sent from the server to train its local model, the edge devices then send its updated weights to the server. It is important to note that the aggregate

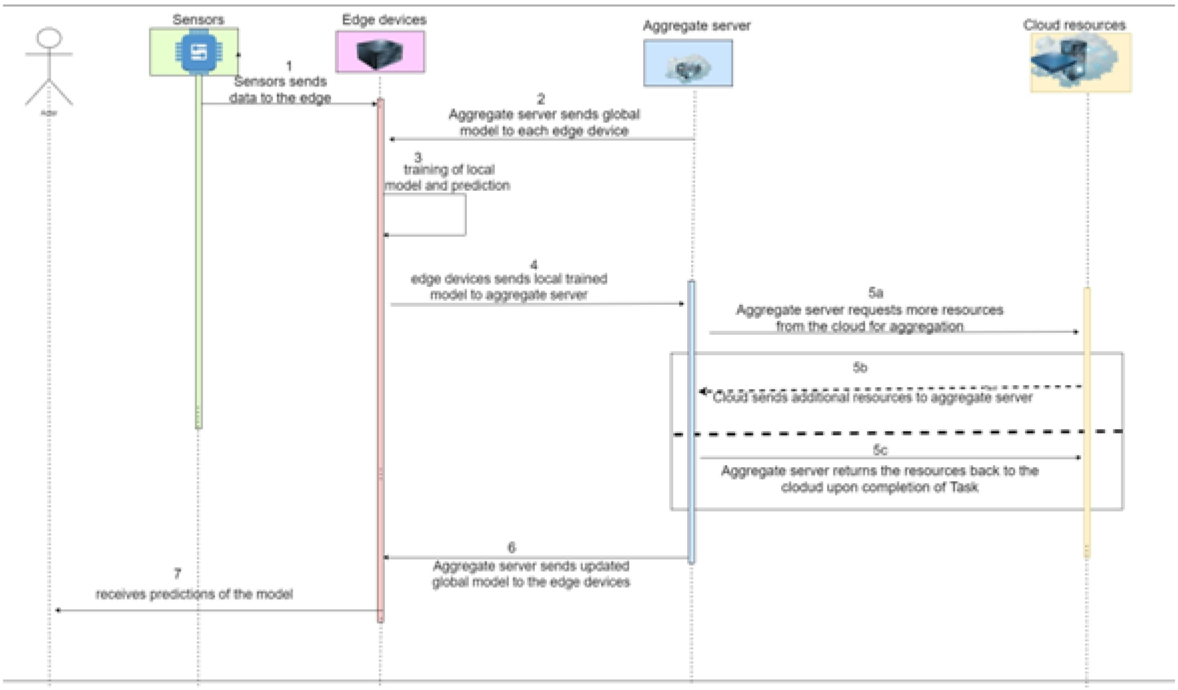

Figure 2: Federated Learning Sequence.

server never sees the raw data of the edge devices throughout the entire process which provides data privacy and security for the data for the entire analysis.

2.2. Gaussian Naïve Base (NB) Classifiers

The authors in [22] discussed that Binary relevance breaks down the multi-class dataset into several independent binary variables such that one variable is in one label. According to [23], the classifier chain Gaussian NB equally disintegrates the multi-class dataset into many independent variables but recognises the dependent variable correlations which is an enhancement over the Binary relevance Gaussian NB model. The authors in [23] discuss that the Label powerset Gaussian NB transform the multi-label dataset into many multi-classes single-label classification problem. The Gaussian Naïve Bayes is implemented from the Naive Bayes theorem.

2.3. Federated Learning

The following steps describe the sequence:

1. Initialisation of the tasks The training task is decided by the server.

The training process and global model hyper-parameters are handled by the server.

The selected participants receive the task and initialise the global model V_p^o

2. update and train the local model.

The edge nodes use their local data and devices to optimize the local model V_p^t

where t represents the recent iteration index.

The purpose of the edge nodes i in the process t is to determine the best variables V_i^t that will decrease the loss function $L\left(V_i^t\right)$

$$V_i^t = \arg\min L\left(V_i^t\right) \text{..........................} (2)$$

	Algorithm 1: Federated Averaging Algorithm [16]
	The Learning rate is η
	The number of local epochs is e
	Locally reduced batch (mini-batch) = S
	Number of edge nodes in each iteration = c
	Global model V_p^o
1.	The participants are represented by i
2.	Local Training V_i
3.	Divide local dataset G_i to small mini-batches, and place in set G_i
4.	s which is part of a set S_i
5.	for every local epoch h, from i to e do
6.	for every $s\epsilon$ S_i do where (η= learning rate and δ=gradient of L on S)
7.	end for
8.	end for
9.	[server]
10.	set V_p^o
11.	for iteration t from 1 to t do
12.	arbitrarily select a subset Y_t of C edge nodes from N
13.	for each edge node i ϵY_t similarly do
14.	V_i^{t+1} local training (i, V_p^t)
15.	end for
16.	aggregating $V_p^t = \frac{1}{\sum_{i\epsilon N} D_i} \sum_{i=1}^{N} D_i V_i^t$
17.	end for

$$V_i^t$$

The server receives the updated local model parameters.
3. Global model accumulation and modification.
Local models are aggregated which are from the edge nodes to the server, the edge nodes receive the modified global model. V_p^{t+1}
$L\left(V_p^t\right)$ is the global loss function, minimised by the server.

$$L\left(V_p^t\right) = \frac{1}{N}\sum_{i}^{N} L\left(V_i^t\right) \text{.........................(3)}$$

The global loss function converges after many repetitions of steps 2 to 3 (state which additional iterations do not enhance the model) FL training using learning rate=0.01, optimizer=SGD.

3. Methodology and Experimental Set-up

The data used for this research include climatic features namely temperature, humidity, the potential of hydrogen(pH), and rainfall which are the independent variables, and the labels are rice, maize, and chickpea. The classes in the dataset namely chickpea, rice, and maize are equally distributed. This implies that the dataset is balanced. The dataset has been split into 80% for training and 20% for testing using the sci-kit learn library [24]. Each federated node has the same labels and attributes since we are exploring homogeneous edge nodes where all the edge nodes manage data with the same attributes and features. The research experiment aims to investigate the prediction of a particular crop from a class of crops using climatic data as the independent features from the dataset, while the crop types are the labels from the dataset. This was achieved using a modified federated averaging algorithm model. The Syft library is used in a decentralised platform where the edge nodes' data reside at the edge nodes and the data scientist remotely trains the dataset without seeing the data [24], this research uses the Syft library in the duet platform in our testbed. The Testbed has been set up using a Linux machine, the data scientist and the data owner have been able to interact via the duet platform, and the Data owner is the custodian of the data. First, the data owner establishes the connection using the duet server and waits for the Data scientist to connect to the data owner via the duet server, once a connection was established, the data owner(edge device) then proceeds to train its dataset and sends its local updated weights to the aggregate server or data scientist, the updated global model is then sent back to the edge devices for a repeat iteration and this process continues until the model converges. An emulation of the network was set up using the GNS3 tool, to test the Federated Learning model for a smart farming dataset, climatic data with independent variables such as temperature, humidity, pH, and rainfall were used as the independent variable while three crops namely rice, maize, chickpea were considered as the dependent variable and the results shown in tables 1-4 were obtained from the experiment.

4. Results and Discussion

The dataset with independent variables of temperature, humidity, pH, and rainfall and dependent variables of rice, maize, and chickpea has been passed into the Binary Relevance (Gaussian NB), Classifier chain (Gaussian NB) and Label Powerset (Gaussian NB) model in the test bed setup within the Jupyter Notebook and the following results have been obtained as shown in figure 3, 4, 5 respectively. The Binary Relevance (Gaussian NB), classifier chain (Gaussian NB) and Label power (Gaussian NB) produced an accuracy of 60%, 60%, and 55% respectively from the training. The

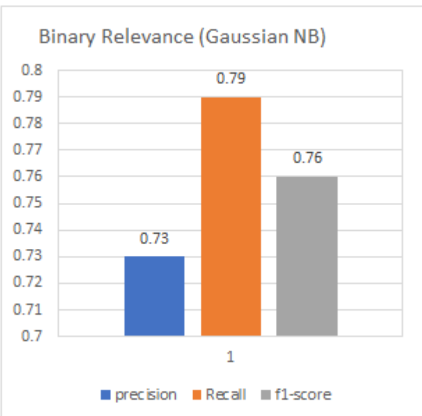

Figure 3: Binary Relevance (Gaussian NB)

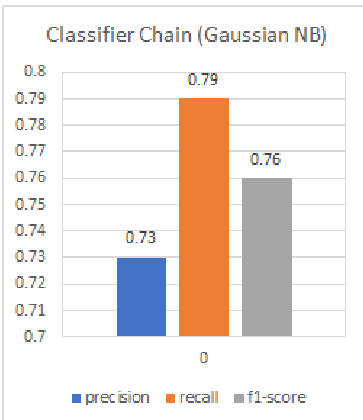

Figure 4: Classifier Chain (Gaussian NB)

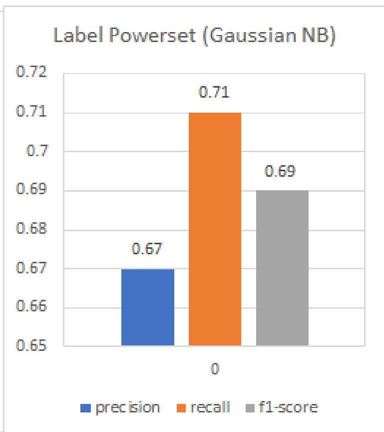

Figure 5: Label Powerset (Gaussian NB)

Binary Relevance and classifier chain Gaussian Naïve Bayes model has been used to evaluate the multi-labelled dataset. Figures 3 and 4 indicate the results obtained from using the Binary Relevance Gaussian NB and Classifier Chain Gaussian NB model in both evaluations, a Harmonic mean of 0.76 and Accuracy of 60% has been obtained. The Binary relevance Gaussian NB and Classifier chain Gaussian NB has been able to use the sample averages of each instance of the multi-labelled dataset to produce a Harmonic mean of 0.76 and both models were able to match 60% of the predicted multi-labelled variables to the original labels of the dataset. The Label Powerset Gaussian NB model has produced an accuracy of 55% as shown in Figure 5. The F1-score of 0.69 has been achieved by the model showing that the ratio of the product of the precision and recall to the sum of the precision and the recall values from the model during evaluation is 0.69. The model takes into account the sample average since the dataset considered is a multi-label and each of the sample averages for each instance is used during evaluation to produce the harmonic mean of the model. Tables 1 – 4 show the results obtained from using the federated learning models to predict the crop type using climatic parameters as independent variables and crops as labels. The

Table 1: FL training using learning rate=0.001, optimizer=SGD

	Precision	**recall**	**f1-score**	**support**
0	0.40	0.60	0.48	0.10
1	0.70	0.10	0.80	0.10
2	0	0	0	0.10
Accuracy			0.23	0.30
macro average	0.16	0.23	0.19	0.30
weighted average	0.82	0.23	0.19	0.30

Table 2: FL training using learning rate=0.01, optimizer=SGD

	Precision	**recall**	**f1-score**	**support**
0	0	0.62	0.77	0.10
1	1	0.30	0.46	0.10
2	0.91	1	0.95	0.10
Accuracy			0.77	0.30
macro average	0.84	0.77	0.73	0.30
weighted average	0.84	0.77	0.73	0.30

model hyper-parameters have been tuned to obtain various results, the learning rate hyper-parameters range from zero (0) to One (1), and different values of learning rates between zero(0) and one(1) have been considered for hyper tuning of the models, more so different optimizers such SGD and Adam has been considered based on previous research by [19]. Using an SGD optimizer, a learning rate of 0.001, and a Computational time of 0.00013 seconds have been obtained during the training of the model. An Accuracy of 23% has been obtained while the predicted crop was rice, implying the model made high errors since its loss values are also high as can be seen in Figure 6. It can be inferred that using the SGD optimizer and a learning rate of 0.001 only 23% of the predicted labels have been matched with the original labels in the dataset after the training which indicates the SGD optimizer at this learning rate produced a poor

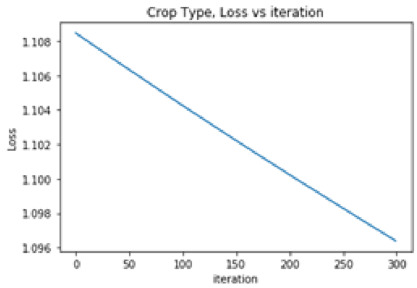

Figure 6: Loss using SGD optimizer, Learning rate=0.001

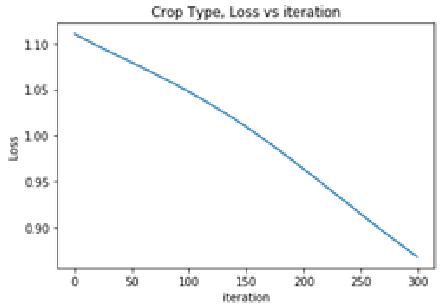

Figure 7: Loss using SGD optimizer, Learning rate=0.01

accuracy and failed to match the predicted classes with the original labels. Figure 7 shows further hyper tuning using the Federated Learning Model using a Learning rate of 0.01, SGD optimizer, the model is converging very poorly due to over-fitting of the model. it produced an accuracy of 77% indicating it has been able to match only 77% of the predicted crop label to the original crop-dependent variables

The Federated learning model produced a precision value of 0.40 using the SGD optimizer and a Learning rate of 0.001. This is the ratio of the correctly predicted positive labels to the sum of the correctly predicted positive labels and the incorrectly predicted positive labels. Upon further evaluation where the model has considered the ratio of the correctly predicted positive labels to the sum of the correctly predicted positive labels and the incorrectly predicted negative labels giving a recall value of 0.60 which can be referred to as the recall value. Comparing the precision and recall values from the federated learning model, an F1 score of 0.48, which is the Harmonic mean, that's the reciprocal of the arithmetic mean has been produced which is a poor performance of the SGD optimizer function, as shown in Tables 1. It can be inferred that the SGD optimizer function with a learning rate of 0.001 converged poorly and extremely slowly to its local minima as shown in Figure 6. Further hyper-tuning of the model parameters has been conducted with the SGD optimizer but with a different learning rate value of 0.01. The results in Table 2 indicate that only 77% of the predicted labels matched the original labels of the classes of chickpea, rice and maize. The model has failed to produce a value for the evaluation of the ratio of the true positive of the predicted labels to the sum of the true positive predicted labels and incorrectly predicted positive labels, this indicates the poor performance of the model using the SGD optimizer and learning rate values of 0.01. The evaluation of the ratio of the true positive of the predicted labels to the sum of the true positive predicted labels and incorrectly predicted negative labels has produced a recall value of 0.60. Taking the ratio of the precision and the recall for the SGD with a learning rate of 0.01, a Harmonic mean (F1-score) of 0.77 has been obtained which is a better performance than the initial learning rate considered earlier. It can be inferred that the federated learning model is converging to its local minima much faster, which is a better value when compared with the results from Table 1 but its performance is unable to give a precision value. From Table 3 a different optimizer function namely the Adam optimizer is considered for the hyper-tuning of the model, the Adam optimizer combines the Adagrad and RMSProp algorithms for its evaluation to give a better evaluation during training. The predicted class has matched the original values

Table 3: FL training using learning rate=0.001, optimizer=Adam

	Precision	**recall**	**f1-score**	**support**
0	0.83	1	0.91	0.10
1	1	0.70	0.82	0.10
2	0.91	1	0.95	0.10
Accuracy			0.90	0.30
macro average	0.91	0.90	0.90	0.30
weighted average	0.91	0.90	0.90	0.30

Table 4: FL training using learning rate=0.01, optimizer=Adam

	Precision	**recall**	**f1-score**	**support**
0	0.83	1	0.91	0.10
1	0.67	0.40	0.50	0.10
2	0.67	0.80	0.73	0.10
Accuracy			0.73	0.30
macro average	0.72	0.73	0.71	0.30
weighted average	0.72	0.73	0.71	0.30

with a percentage of 90% which indicate a good performance of the accuracy metric. The ratio of the correctly predicted positive labels to the sum of the correctly predicted positive labels and the incorrectly predicted positive labels gave a value of 0.83 precision value as shown in Table 3. To further verify the Adam optimizer performance using a learning rate of 0.001, the ratio of the precision and recall values are taken which produce a Harmonic mean (F1-score) of 0.91 from the model evaluation. It can be inferred that the model has converged very fast which enabled it to reach its local minima, thereby improving its performance with a 0.91 harmonic mean (F1-score) value. Further analysis using the Adam optimizer with a learning rate of 0.01, the hyper-tuning of the model, the predicted class has a match with the original values with a percentage of 90% which indicate a good performance of accuracy metric as shown in Table 4. The ratio of the correctly predicted positive labels to the sum of the correctly predicted positive labels and the incorrectly predicted positive labels gives a value of 0.73 precision value. To further verify the Adam optimizer using a learning rate of 0.01, the ratio of the precision and recall values are taken which produce a Harmonic mean (F1-score) of 0.91 from the model evaluation. It can be inferred that the model dropped on its accuracy metric from the previous value using the 0.001 learning rate when a learning rate of 0.01 is considered but has been able to maintain the F1 score. It can be inferred that the model using the Adam optimizer has been able to converge to a local minimum, considering all the true and false positives, and true & false negatives to give a high harmonic mean (F1-score) at a higher learning rate of 0.01. The dataset contained three (3) classes in the dependent variables, during each hyper-tuning with different optimizer functions and learning rate parameters, it has been observed that chickpea was the predicted crop, indicating the federated learning model without seeing the raw dataset has been able to match a higher percentage of the predicted crop with its original values. Figures 6 and 7 show the loss value decreasing during the training of the model using stochastic gradient descent (SGD) optimizer,

Figure 8: Loss using optimizer=Adam, Learning rate=0.001

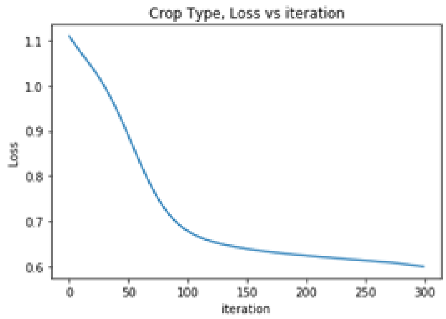

Figure 9: Loss using Adam optimizer, Learning rate=0.01

with a learning rate (LR) of 0.001 and 0.01 respectively.

From the results obtained, as shown in Figure 6, a minimum Loss of 1.096 has been obtained from the evaluation of the model, Figure 7 has produced a minimum loss value of 0.7, while Figure 9 depicts that a minimum loss value of 0.6 and the loss started to converge appreciably after 100 iterations. However, from Figure 8, the loss has started to converge appreciably after 20 iterations and eventually converge at a Loss value of 0.1 which is a better improvement compared with the other initial learning rate of 0.001, 0.01 for SGD optimizer and a learning rate of 0.001 for the Adam optimizer. It can be inferred that with the learning rate of 0.001 using the Adam optimizer, the federate learning model has been able to reach its local minima, although its training time at this learning rate has been increased as shown in Figure 8. However, in Figure 9 its training iteration is over 200, this implies the model has begun to learn the noise in the dataset and it causes over-fitting and generalising poorly. This research results confirm the efficiency of the Adam optimizer from the hyper-tuning of the parameters of the Federated Learning model to a smart farm dataset, it can be inferred that the Adam optimizer converges better than the SGD optimizer. This confirms that federated learning models also reach their local minima at low learning rates and use high training time to converge. The dataset used for this experiment was obtained from [20].

5. Conclusion

A dataset obtained from [15] has been used for this research to determine the performance of the Federated Averaging algorithm within a smart farming scenario. It has been observed that climatic parameters can be considered as independent features and crop types as dependent features, upon training the dataset with the adjusted model, it has been observed that the Adam optimizer has enabled the model to reach its local minima while considering the true and false positive predicted label classes, true and false negatives predicted dependent variables to achieve a harmonic mean (F1-score) of 0.91. It can be inferred from Table 1- 4, which depicts the various Harmonic mean values obtained from the evaluation of the multi-labelled dataset with temperature, humidity, pH, and rainfall as independent variables, with rice, maize and chickpea as labels, using the binary relevance Gaussian NB, Classifier chain Gaussian NB, Label Powerset Gaussian NB and the Federated averaging models that, the optimal harmonic mean has been produced by the Federated averaging model with a value of

0.91 which is the decentralised model where the raw dataset has not been shared, unlike the centralised network where the raw dataset has been shared in the Gaussian NB models. Academic researchers can consider this work results to take decisions on smart farming within a Federated learning platform.

6. Future works

The Swin Transformer can be considered for evaluation of the climatic parameters to predict the crop type. It will be novel research to use the Federated split learning model to predict the crop types using the climatic parameters as independent variables and the crop types as your dependent variables.

References

[1] Juyoung Park, Aekyung Moon, Eunryung Lee (2021), Understanding IoT climate Data based Predictive Model for Outdoor Smart Farm, 2021 International Conference on Information and Communication Technology Convergence (ICTC) — 978-1-6654-2383-0/21/$31.00 ©2021 IEEE — DOI: 10.1109/ICTC52510.2021.962097

[2] Md Toufiqur Rahman, Sakib Mahmud, Yue Li, Md Abdur Rahman (2021), IoT based smart farming system to reduce manpower, wastage of time & natural resources in both traditional & urban mega farming, 2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE) — 978-1-6654-1596-5/21/$31.00 ©2021 IEEE — DOI: 10.1109/AEMCSE51986.2021.00241

[3] Qinbin Li, Zeyi Wen, Zhaomin Wu, Sixu Hu, Naibo Wang, Yuan Li, Xu Liu, Bingsheng He (2021), Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection, arXiv:1907.09693v6 [cs.LG], 1 Jul 2021

[4] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, Virginia Smith (2019), Federated Learning: Challenges, Methods, and Future Directions, arXiv 1908.07873v1 [cs.LG] 21 Aug 2019 28

[5] Shiqiang Wang, Tiffany Tuor, Theodoros Salonidis, Kin K. Leung, Christian Makaya, Ting He, Kevin Chan (2019), Adaptive Federated Learning in Resource-Constrained Edge Computing Systems, arXiv:1804.05271v3 [cs.DC] 17 Feb 2019

[6] Shaoxiong Ji, Shirui Pany, GuodongLongz, Xue Li, Jing Jiangz, Zi Huang (2019), Learning Private Neural Language Modeling with Attentive Aggregation, IJCNN 2019. International Joint Conference on Neural Networks. Budapest, Hungary. 14-19 July 2019

[7] Vagisha, E. Rajesh, S. Basheer and K. Baskar, (2023) Hydroponics Soilless Smart Farming in Improving Productivity of Crop Using Intelligent Smart Systems, 2023 3rd International Conference on Innovative Practices in Technology and Management (ICIPTM), Uttar Pradesh, India, 2023, pp. 1-6, doi: 10.1109/ICIPTM57143.2023.10117747.

[8] W. Lai and Q. Yan, (2022) Federated Learning for Detecting COVID-19 in Chest CT Images: A Lightweight Federated Learning Approach, 2022 4th International Conference on Frontiers Technology of Information and Computer (ICFTIC), Qingdao, China, 2022, pp. 146-149, doi: 10.1109/ICFTIC57696.2022.10075165.

[9] H. Sifaou and G. Y. Li,(2022), Robust Federated Learning via Over-the-Air Computation, 2022 IEEE 32nd International Workshop on Machine Learning for Signal Processing (MLSP), Xi'an, China, 2022, pp. 1-6, doi: 10.1109/MLSP55214.2022.9943401.

[10] K. I. -K. Wang, X. Ye and K. Sakurai,(2022), Federated Learning with Clustering-Based Participant Selection for IoT Applications, 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 2022, pp. 6830-6831, doi: 10.1109/BigData55660.2022.10020575.

[11] Catalfamo Alessio, Carnevale Lorenzo, Galletta Antonino, Martella Francesco, Celesti Antonio, Fazio Maria and Villari Massimo, (2022), Scaling Data Analysis Services in an Edge-based Federated Learning Environment," 2022 IEEE/ACM 15th International Conference on Utility and Cloud Computing (UCC), Vancouver, WA, USA, 2022, pp. 167-172, doi: 10.1109/UCC56403.2022.00030.

[12] Jakub Konecny, H. Brendan McMahan, Felix X. Yu, Ananda Theertha Suresh Dave Bacon, Peter Richtarik (2017), Federated Learning Strategies for improving communication efficiency, arXiv 1610.05492v2 [cs.LG], 3, Oct 2017

[13] Jin-Hyun Ahn, Osvaldo Simeone, and Joonhyuk Kang (2020), Cooperative learning via federated distillation over fading channels, 978-1-5090-6631-5/20/$31.00 c 2020 IEEE

[14] Kai Yang, Tao Jiang, Yuanming Shi, and Zhi Ding (2020), Federated Learning via Over-the-Air Computation, IEEE Transactions on wireless communications, Vol. 19, No. 3, March 2020 10. Sumudu Samarakoon, Mehdi Bennis, Walid Saad, and Merouane Debbah (2018), Federated Learning for Ultra-Reliable Low-LatencyV2V Communications, 978-1-5386-4727-1/18/$31.00 c 2018 IEEE

[15] Guan-Ying Huang and Ching-Hung Lee (2021), Federated Learning Architecture for Bearing Fault Diagnosis, 2021 International Conference on System Science and Engineering (ICSSE) — 978-1-6654-4848-2/21/$31.00 ©2021 IEEE — DOI:10.1109/ICSSE52999.2021.9538492

[16] David Leroy, Alice Coucke, Thibaut Lavril, Thibault Gisselbrecht and Joseph Dureau (2019), Federated Learning for keyword spotting, 978-1-5386-4658-8/18/$31.00 c 2019 IEEE, ICASSP 2019

[17] Dequan Li, Yuheng Zhang, Yuejin Zhou (2021), Fast Distributed Stochastic Nesterov Gradient Descent Algorithm for Image Classification, 2021 China Automation Congress (CAC) — 978-1-6654-2647-3/21/$31.00 ©2021 IEEE — DOI: 10.1109/CAC53003.2021.9727635

[18] Tao Sun, Linbo Qiao, Qing Liao, and Dongsheng Li (2021), Novel Convergence Results of Adaptive Stochastic Gradient Descents, IEEE Transactions on image processing, Vol. 30, 2021

[19] Diederik P. Kingma and Jimmy Lei Ba (2017), Adam: A method for stochastic optimization, arXiv:1412.6980v9 [cs.LG], 30 Jan 2017

[20] Arthava Ingle (2020), https://www.kaggle.com/atharvaingle/crop-recommendation-dataset

[21] Ming Qiu, Yiru Zhang, Tianqi Ma, Qingfeng Wu, and Fanzhu Jin (2020), Convolutional-neural-network-based Multilabel Text Classification for Automatic Discrimination of Legal Documents Sensors and Materials, Vol. 32, No. 8 (2020) 2659–2672 MYU Tokyo, https://doi.org/10.18494/SAM.2020.2794, ISSN 0914-4935 © MYU K.K.A.N.M. JuBaer.

[22] Abu Sayem and Md. Ashikur Rahman (2019), Bangla Toxic Comment Classification (Machine Learning and Deep Learning Approach), Proceedings of the SMART–2019, IEEE Conference ID: 46866, 8th International Conference on System Modelling & Advancement in Research Trends, 22nd–23rd November 2019, College of Computing Sciences & Information Technology, Teerthanker Mahaveer University, Moradabad, India.

[23] Liao Xiaoqun, Cao Nanlan, Ma Li, Kang Xiaofan (2019), Research on Short-term Load Forecasting Using XGBoost Based on Similar Days, 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), 978-1-7281-1307-4/19/$31.00 ©2019 IEEE DOI: 10.1109/ICITBS.2019.00167

[24] Adam James Hall, MadhavaJay, Tudor Cebere, Bogdan Cebere, Koen Lennart vander Veen, George Muraru,Tongye Xu, Patrick Cason, William Abramson, Ayoub Benaissa, Chnimay Shah, AlanAboudib, Th´eoRyffel, Kritika Prakash, Tom Titcombe, Varun Kumar Khare, Maddie Shang, Ionesio Junior, Animesh Gupta, Jason Paumier, Nahua Kang, Vova Manannikov, and AndrewTrask, SYFT0.5:A platform for universally deployable structured transparency , ICLR2021-Workshop on Distributed and Private Machine Learning (DPML), arXiv:2104.12385v2 [cs.LG] 27 Apr 2021