International Journal of Advanced Trends in Computer Science and Engineering

A Comparative Study of Various Machine Learning Algorithms in Fog Computing

Urooj Yousuf Khan[1], Muhammad Mansoor Alam[2]
[1]Institute of Business Management, Pakistan, urooj.yousuf@iobm.edu.pk
[2]Institute of Business Management, Pakistan, malam@iobm.edu.pk

ABSTRACT

Internet of Things is the reality of the future. It encompasses a vast prototype. The applications are limitless and the research potential is huge. Several unanswered questions still stand in the way. One of these challenging questions is the Network layer implementation of IoT. Cloud offers one perspective solution. Cloud services face the down side of connectivity and latency. Fog computing, as a complementing technology for Cloud answers the issues of latency and connectivity in the Cloud. Being novel idea in itself, Fog computing offers promising future growth. A major concept in Fog networks is its ability to learn and adopt to changing environment. This ability of the Fog devices to self-learn and improve itself is the core of machine-learning. Hence, machine-learning algorithms occupy central place in visualizing robust and effective Fog networks. This paper presents a comparative study of various machine-learning algorithms and their applications in the Fog network. The paper also projects the undiscovered research paradigms and future directions

Key words: Cloud Computing, Fog Computing, Internet of Things, Machine Learning

1. INTRODUCTION

Fog as a probable platform for envisioning Internet of Things (IoT) is a viable and sustainable idea. Visualization of being connected everywhere is largely supported by the Cloud and Fog computing[1] . Another way to realize IoT are the cyber-physical systems [2]. A prominent feature of any smart cyber-physical network is its ability to learn from the environment and the presence of feedback loops. These feedback loops ensure that data analysis and study take central stage [2]. In the perspective of IoT, this data analysis takes place at the Cloud. Data collection and immediate processing takes place at the Fog node. For long- term data analysis and processing, a feedback system and a mathematical model is needed. Embedding the ability of self-learning in the Fog nodes would serve similar purpose. There are many arguments to support idea of smart nodes. These include deep knowledge from data inspection, support

in the informed decision-making process and ultimate integration of the smaller systems into lager one [2]. Thus, this system of intelligent Fog nodes and Cloud provides a promising platform to realize IoT.

If envisioned completely, there are huge undiscovered areas where IoT can bring about revolutionary changes[3]. These applications have a huge influence on human lives, particularly in the health monitoring systems [4]. The paper is organized as follows: Section 2 covers the literature review. Section 3 covers the Fog paradigm in machine learning. Section 4 covers the application areas where machine learning can enhance the functionality of Fog Networks. Section 5 points to newer research directions and Section 6 concludes the paper. References are covered in the last section.

2. LITERATURE REVIEW

2.1 Cloud and Fog

Cloud computing has emerged as one of the most dominant technologies. It is immensely capable of storing, analyzing and managing huge amount of data. The focus of Cloud computing is largely on centralized and remote data centers [2]. This design comes with various shortcomings. These include security, reliability and latency[2],[1]. Fog, as a compliment to the existing Cloud network answers these drawbacks[1]. It permits better resource utilization and lower latency. Localized data processing enhances reliability and security.

Fog is a term coined by Cisco[3],[1] in 2012. It is an extension of the Cloud. It has many similar equivalents such as edge computing, mobile computing, grid computing or cloudlets[1]. Concisely, it is a localized data center with computational capabilities. The aim is to provide control and communication to the nodes in real time. It answers many downsides of Cloud such as security, bandwidth consumption and latency. These challenges are expected as the networks expand to 5G. The introduction of Cloud and Fog can further envision and enhance IoT and artificial intelligence [3].

In 2015, OpenFog Consortium was founded. It is a group of distinguished companies namely CISCO, Dell and academia such as Princeton University. A reference architecture was released in 2017. It is built upon guiding pillars of agility,

autonomy, availability, hierarchy programmability, serviceability, reliability, openness and security [5]. There are enormous potentials for the Fog as a complement to the Cloud. These include great geographical distribution, task offloading and latency minimization[5].

Utilizing the smart interconnected Fog-Cloud network to realize IoT offers great advantages. This novel idea has far reaching influence in human lives. A study proposed a health monitoring system based on IoT and machine learning [4]. The study used three-tier architecture of IoT to monitor significant parameters for heart diseases [2]. According to another study, an interesting analogy exists between human brain and these self-learning Fog networks[6]. Plasticity is the capability of human brain to learn from its environment and adapt to the changes. Moreover, the parallel execution of concurrent tasks, selection of the best-fit algorithm for a given set of constraints and job scheduling are tasks of the human brain. Human brain presents one of the most powerful and efficient processing systems[6].

Technically, Fog as a communicating layer between Devices and the Cloud is a vastly supported idea. This multi layered architecture chiefly includes User-Equipment layer, Network layer and Data Processing layer [7]. There are different implementation details and challenges for each of these layers. Fog nodes collect user data from User-Equipment layer, process it locally at the Network Layer and route it to Cloud at the Data Processing Layer [7]. Conferring to authors, this hierarchical network of User devices, Fog nodes and Cloud has certain limitations[8]. Major challenges in this layered implementation for IoT include security, trust and privacy [8]. Inclusion of Fog nodes has majorly redefined the bounds. These Fog nodes now provide localized solutions with better bandwidth conservation, decentralized processing and real-time analysis [8].

Another example of including Fog nodes to enhance productivity exists in literature[9] . According to the study conduct, one such scenario is discussed bysimulating the data stream in Fog instead of routing it to the Cloud. The results showed that the proposed model achieved 97% accuracy with an end-to-end delay of 180ms[9].

2.2 Machine Learning

Machine learning aims at improving the performance of the computers by learning through familiarity. It enables computers to improve their response to problems with experience. This response can be applied to solving similar problems in the future. Classically, there are following three major divisions of machine-learning algorithms: supervised, unsupervised and reinforcement learning. Supervised learning includes mapping. It charts input values to predefined output. These problems majorly fall under two sub-classes: classification and regression. Unsupervised learning implies finding data patterns in an unsupervised manner. Reinforcement learning largely comprises of a series of feedback loops for reward and penalty. In this environment, the intelligent agent (learner) is an independent decision-maker. This decision maker takes decisions based on the awareness of the environment [2].

Incorporating machine-learning algorithms in Fog nodes to make them more intelligent implies introduction of "plasticity" within these nodes[6].

Many studies suggest that machine learning is resolving key issues within the Fog Computing architecture. These include data aggregation, computational overhead and communication efficiency among many others [10]. This study suggests a method for data aggregation and analysis in Fog nodes using Machine Learning[10]. A process-monitoring and prediction system based on Machine learning offers an efficient alternate to existing designs [11]. This system utilizes the Random forest algorithm to make scalable, high performance predictive model[11] . The proof of concept provides a stronger foundation to the concept[11]. The task of quickly learning from the surroundings and utilizing the limited bandwidth is further elaborated with the container designs [12]. The study utilizes machine learning algorithms such as Q-learning and deep learning to handle large-scale migrations in Fog nodes [12]. The results concluded that deploying machine learning techniques such as Deep learning significantly reduced the delay, power consumption and migration cost.

Another significant area where machine learning is changing the paradigm is the security and safety in Fog Computing system[13].The heterogeneous nature of the Fog nodes and their scalability leads to the problems of authentication and security within the local network. Deploying machine learning techniques for detecting anomalies and mitigating attacks is discussed here [13].

Table 1 presents machine-learning solutions for numerous IoT security threats[13].

Table 1: Machine learning solutions for numerous IoT security threats[13]

Use-Case	Machine Learning Algorithm
Anomaly Detection	Naïve Bayes ANN
Finding Infrequent Data points	Support Vector Machine Random Forest PCA Naïve Bayes KNN
Forecast Values and Class	Linear Regression Support vector Regression CART FFNN
Feature Extraction	PCA CCA
Malware Detection	SVM Random Forest
Intrusion Detection	PCA Naïve Bayes
Unearthing of pattern	K-Means DB-SCAN

The security threats in the smart city come naturally since it is a fully connected entity. These threats can have great impact on human life. Another machine learning algorithm is applied to detect cyberattacks in the Smart City [14]. AD-IoT

is a proposed system for detecting anomalies in the system using Machine learning. This model also deploys Random Forest technique for anomaly detection. The paper also suggests enhancements such as deploying Conventional Neural Networks (CNN) to control traffic activities in the city. This effective algorithm gains the highest classification accuracy of 99.34% with lowest false positive rate[14].

Another interesting case study for deploying machine learning algorithms is found in BEMS [15]. In this paper, the authors propose a self-learning algorithm using Neural Networks. The smart buildings automatically control the temperature, air quality and the humidity within the premises. Deploying an effective machine learning algorithm such as Neural Network has enabled examination and higher accuracy rate.it has also led to reduced energy consumption and increase of comfort level for the residents [15].

An application of task offloading in Fog Networks using classification and regression tree algorithm is also found in literature[16]. This study suggests that usage of classification algorithm has reduced power consumption of CPU, RAM and Wi-Fi by 44%,71% and 55% respectively[16] . Usage of machine learning in vehicular computing and connected vehicles is another significant paradigm [17]. The effective usage of machine learning methods aims at improving the mobility and connectivity. The loss of connectivity when transitioning from one fog node to the other can be minimized using recurrent neural networks (RNN)[17] . The selection of optimal node and its effective cost were calculated given the time and location. This optimal node minimized the handover cost associated with loss of connectivity [17]. It is expected that more than 9 billion people will live on the planet by 2050 and about two thirds of them will live in cities [18]. An important implementation of the Fog computing lies in Smart cities. Fog computing enables huge data generated by the nodes to be processed locally. Incorporating machine intelligence in the system comes with challenges such as the capability to handle such diverse data resources and physical constraints imposed by device limitations [19].

2.3 Fog Architecture

The precise location and platform selection of the Fog nodes depend on the application scenario and required number of hosts. There are key differences between Cloud and Fog. The major difference lies in utilization of resources. Fog is a resource-constrained environment with stringent time requirement. Cloud, on the other hand, virtually has unlimited storage and processing capability. Moreover, peer-to-peer communication is much faster in Cloud due to high-speed connectivity. It does not happen in Fog where the networks vary quickly and communication is majorly wireless. Vertical communication, however, takes another form. In Cloud, this communication is slower while accessing remote devices. Cloud computing is a stable, market-oriented paradigm while Fog is still an emerging concept [1].

Table 2 below identifies the key differences between Cloud and Fog.

Table 2: Differences between Cloud and Fog [20]

Feature	Description
Heterogeneity	Located at the edge of the network with rich, heterogenous end-user support
Capability	Supports a wide range of industrial applications deo to instant response capability
Storage and Services	Has its own computing, storage and networking services
Operation Areas	Operates locally single hop from a device to a fog node)
Platform	Has a highly virtualized platform
Additional Features	Portable deployment in terms of hardware and software

Table 3: Computational aspects for Cloud and Fog[20]

Aspects	Fog Computing vs Cloud Computing
Resources	Relatively fewer computing resources in Fog
Capability	Both can process data generated from a diverse set of resources
Distribution Strategies	Based on geographical location, both can distribute data densely or sparsely
Connectivity	Both support machine-to-machine communication and wireless connectivity
Flexibility	Fog can be installed on low specification devices such as switches and Internet protocol cameras
Usability	One of the primary uses is currently for mobile and portable devices

Table 3 above provides a clear picture for different computational aspects of Cloud and Fog. There is a separate paradigm for Fog devices. These Fog devices can be further classified as mission-critical, real-time, interactive, conversational, streaming, CPU-bound and Best effort[21]. Each class clearly specifies the significance of Fog computing in its own terms[21].

Table 4 below signifies the architectural differences between Cloud and Fog.

Table 4: Differences between Fog and Cloud [5]

Features	Cloud Computing	Fog Computing
Architecture	Centralized	Decentralized
Access	Fixed and Wireless	Mainly Wireless
Latency	High	Very Low
Mobility	Limited Support	Supported
Operation	Operated by large companies	Often operated by small companies
Server Nodes Number	Few	Very Large
Server nodes location	The Internet	The local network
Size	Very large data centers	Large number of small fog nodes
Scalability	Average	High
Vulnerability	High probability	Very low probability

Figure 1 below shows the Fog architecture and its different classes. Evident from the figure is the placement and application of the Fog nodes. The end devices could be any user-defined input scheme. The Fog layers 1 and 2 provide the layered implementation for varied types of requirements.
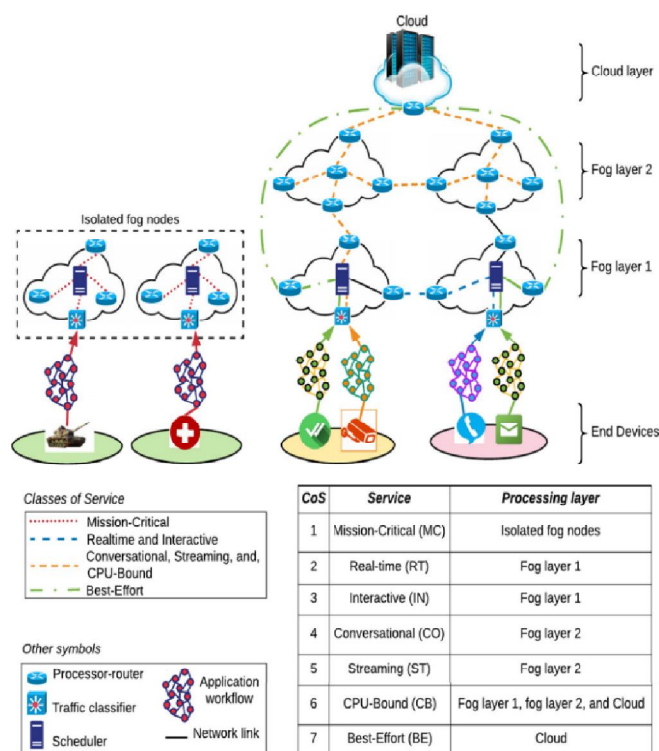

CoS	Service	Processing layer
1	Mission-Critical (MC)	Isolated fog nodes
2	Real-time (RT)	Fog layer 1
3	Interactive (IN)	Fog layer 1
4	Conversational (CO)	Fog layer 2
5	Streaming (ST)	Fog layer 2
6	CPU-Bound (CB)	Fog layer 1, fog layer 2, and Cloud
7	Best-Effort (BE)	Cloud

Figure1: Fog Computing Architecture and Class of Service [21]

Another example of layered approach for implementing CPU-bound, real-time Fog paradigm exists in IoTPulse where Fog devices play major role in health monitoring sector[22] . The system comprised of four layers: sensing layer, network layer, servicing layer and interfacing layer. The sensing layer gathered the patient parameters while the network and servicing layer utilized the Fog networks for decision making. Classification of Fog nodes into heterogenous cellular networks further promotes latency reduction to achieve better Quality of Service (QoS) [23].
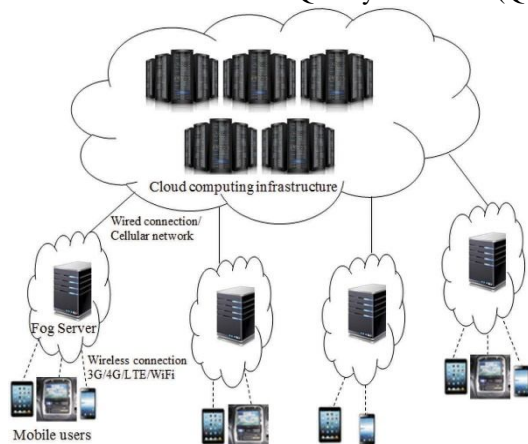

Figure 2: Fog Computing Infrastructure[24]

Figure 2 further explains Cloud-to-Fog continuum. The Fog nodes, placed in user proximity provide continuity along the Things-to-Cloud. Optimal placement of nodes in the Fog network also promotes job scheduling and deploys digital resources at the Edge of the network. It provides a local pool of resources for immediate and optimal utilization[24].

Figure 3 below clearly depicts the placement of Fog Nodes and their functionality. A typical Fog network is identified by a large number of heterogeneous nodes, placed in close proximity to the user or test case. These nodes vary largely from each other in terms of data streams, protocols and networking capabilities. These may include PCs, routers, switches, proxy servers, mobile devices or source inhibited System on Chip (SoC). These heterogeneous nodes communicate with each other via Ethernet or Wi-Fi based networks. These are resource stringent nodes that provide more options for wireless communications. In other words, the great diversity among nodes requires flexible and elastic framework, with real time response. All thesefeatures are intrinsic to Fog. Thus, Fog is becoming a pivotal point in future networks [25].
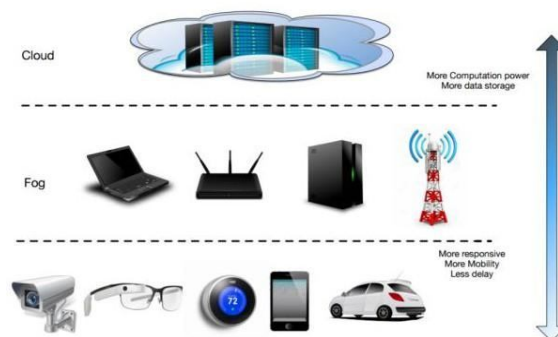

Figure 3: Architecture of Fog Networks[25]

Fog device acts as an intermediary layer that "sits" between Cloud and the heterogeneous ends. The challenges for the Fog nodes include dynamic nature of the input data, the varied data syntax and semantics and minimum latency time. All these scenarios are clearly depicted in the figure above. Moreover, the significance of machine learning can be visualized clearly. A robust self-learning mechanism would enhance the productivity and efficiency of these networks.

Figure 3 complete architecture of Fog. It explains the communication between IoT, Fog and Cloud. The layered architecture of Fog is clearly visible. The Fog layer provides hierarchical continuum along Things-to-Cloud for limited, time-bound data storage. This heterogeneous arrangement provides varied services for different user requirements. These include scenarios such as connected cars, smart grids, intelligent buildings etc. These nodes are essentially a huge collection of geographically smart devices such as routers, switches or even an RFID enabled device [11],[5]. These nodes receive and send data to nearest available nodes horizontally in their network [5].

All figures above identify yet another interesting scenario. It is highlighting the IoT paradigm. The topological arrangement of the Fog devices and the Cloud networks is also enhancing IoT visualization. Here the varied devices

and different physical objects interact and communicate with each other using a common communication platform. Initially the concept was to utilize the IPV4 addressing. However, later it was upgraded to IPV6, thereby making it possible for every device in the planet to be uniquely identifiable[2].

The scope of Fog is not only limited to resolving the bottlenecks faced by Cloud. This paradigm extends far beyond that. It can form core component in emerging IoT, embedded AI and 5G among many others. A considerable advantage of Fog Computing also lies in the fact that it possesses pool of resources that can be utilized as per the user requirements. Moreover, being near to the end users promotes greater cognition and swift data analysis. All these features enhance efficiency. Additionally, bandwidth conservation is ensured, as data does not travel to a central node or server[3] .

On a detailed note, the Fog architecture is denser and more versatile compared to Cloud. Fog computing embeds IoT and business insights gathered from nodes in real-time. It provides distributed and embedded intelligence. This results in scalability and heterogeneity in Fog[10]. Fog servers are dense collection of sensors and are geographically located at closest proximity to the users. These features enhance lower latency, higher mobility and lower deployment cost.

2.4 Machine Learning Associated Algorithms

Below are some of the widely used machine learning algorithms.

A.Decision Trees and Random Forests
 These algorithms belong to the class of algorithms that follow ranking. The ranking method is implemented through tree-like structure. Nodes constitute decision points and leaves constitute predicted outputs. The leaf topology is based on specific, predefined attribute. A major advantage of these algorithm is their understandability and readability [2]. There are two sub branches: classification and regression.

A group of random trees forms the forest. These random trees are developed through bagging. Bagging process divides the dataset of X samples into y smaller, unique datasets. Each of these y datasets trains a different decision tree.

The heuristic followed here is as follows: ''combination of the resulting models should be more robust to overfitting [2].'' The output produced by each tree is combined together by a variety of rules.

B. Artificial Neural Networks
Neural networks are one of the most efficient machine learning algorithms that aim to simulate human brain and underlying communication network [7]. It works by modelling nervous system in living beings. The functional unit of the nervous system is a single neuron. The nervous system comprises of multiple layers of interconnected neurons. These pseudo-neurons assist in conducting tasks such as data mining, pattern recognition and classification [7]. The input layer moves to the next layer using a non-weighted combination to generate transformation values. The process continues until the output layer is reached. These algorithms

are capable to represent complex relationships among the elements of the input set [2]. Another variation of the ANN is in multilayer perceptron (MLP) that deploys nonparametric estimator. Convolutional Neural Networks (CNN) works on the principal of translational invariance. It implies that the value of the output function is unchanged as the input function translates. In other words, if the input translates, the CNN will still be able to detect which class the input belongs. CNN majorly comprises of two parts: feature extractor and a trainable classifier such as SVM. The feature extractor reads and extracts the features and characteristics from the raw data. It comprises of multiple similar stages and layers. Classification is performed by the second part: classifier. Another popular class of ANN is Recurrent Neural Networks (RNN). These neural networks have additionally employed the quality of recurrence due to which they are especially important in data analytics and streaming. Deep neural networks take a step ahead. They utilize stacks of multi-layer neurons to efficiently represent complex relationships among data elements [2]. Swift, parallel and distributed computations are key features of Neural Networks [22].
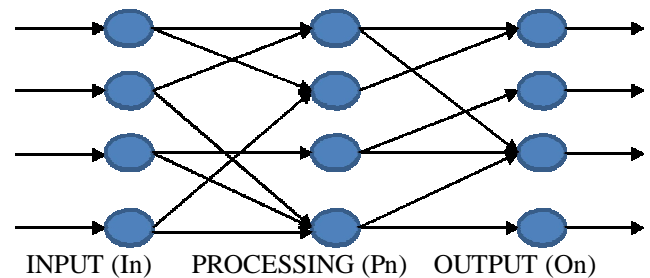Figure 4 below captures the neural networks briefly.

Figure 4. Multilayer Perceptron Neuron Network [22]

C. Support Vector Machines (SVMs)
These machines work under the principle of boundary consideration.
  These supervised learning models segment data by drawing a hyperplane with maximum margin between data points of each class. These machines are scalable and can be cascaded to work on multilevel classification. Due to their flexibility; these SVMs have been extended to solve regression problems as well [2].Moreover, there are notable applications such as Long-Bone X-rays[26].

D. Bayesian Networks
These networks utilize conditional probability to model the system behavior [22]. These networks apply statistical techniques to describe probabilityrelationship among different elements. Bayesian networks indicate the probability of occurrence of an event given that the prerequisite event has already taken place [22]. The method works on directed acyclic graphs. Nodes represent the random variables and directed vertices indicate the influences [2].
Notable variations of Bayesian Networks include Naïve Bayes and Hidden Markov Model. This machine learning algorithm is ideal in modelling uncertain domains [2]. It can be used in supervised and unsupervised learning domains successfully [2].

E. Evolutionary Computations and Genetic Algorithms

This family of algorithms for machine learning revolve around the principles of biological evolution. The guiding principal is ''survival of the fittest [2]''. The incremental learning approach here involves trial-and-error guided by the fitness function. The fitness function measures the goodness or reliability of the output. The fitness function involves working on the principles of natural selection, mutation and cross-overs [2]. This algorithm is applicable in supervised, unsupervised and reinforcement learning [2].

F. Bees-Swarm

This algorithm is also biologically inspired. It imitates two major behaviors in a bee's life: marriage and food foraging [24]. The algorithm begins by a random search of the space to form initial population. As with any other genetic algorithm, this initial population is evaluated by calling the cost function. After the primary filtration, queen is selected as the fittest among N bees. The following 'D' fittest bees are chosen as the drones while the remaining 'W' bees constitute the worker group. The values of 'N','D' and 'W' are predefined by the user. After the selection, the two important events, namely marriage ad food foraging begin. Marriage or reproduction follows typical steps of genetic algorithm namely mutation and crossover to propagate the desired qualities to the next generation. Workers carry out the task of food foraging. It involves finding the food source and recruiting several other workers to collect it. More workers will be recruited if the location is a desirable spot (flowers). Moreover; the best worker is selected for every region to be kept alive for the next population. The cost function is used to select the best worker [9]. The entire algorithm for Bees swarm is given in Figure 5 below. According to the figure, the life cycle of the bees is iterative. It majorly comprises of bees' selection for queen and food search. The genetic algorithm clearly displays the major processes of the life cycle. The cross over and mutation are representing the reproductive cycle. The food search is carried out by the worker bees. The entire life cycle and behavior of the bees is clearly depicted in Figure 5.
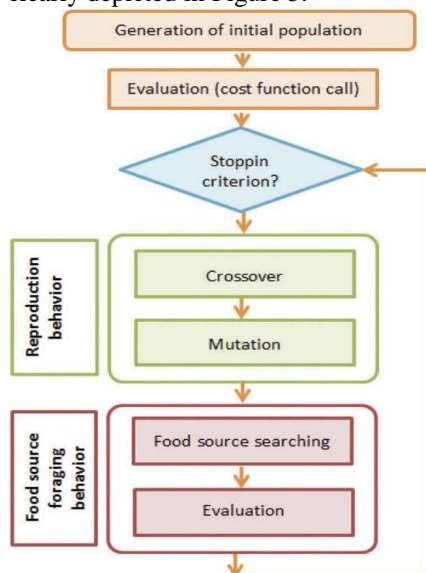

Figure 5. Bees Life Algorithm pseudo-code [24]

G. Clustering

This family of algorithms involve partitioning and grouping elements. The input data set is divided into smaller datasets called clusters. The goal is usually to maximize the similarities among the objects within the same cluster and to minimize the distance between the center of the cluster and individual object. Moreover, the classification of data also involves finding the difference among various clusters. A popular algorithm of this classis K-Means clustering where k denotes the number of clusters [2].

H. Self-organizing Map (SOM)

This algorithm is best suited for topographical distribution. This method is largely applied to clustering problems. SOM implies using finite data models to represent data distribution. A grid is used to represent the arrangement of nodes. The adjacent nodes in the grid represent similar models whereas less similar models are located far away from each other [2]. *Q learning*: It is an incremental, reinforcement algorithm. The agents learn how to act optimally in controlled Markovian environments. It involves techniques for dynamic programming thereby utilizing limited computational space. By successively improving the quality of specified actions, the algorithm reaches the desired results [2].

I.KNN

This specialized, lazy algorithm works by finding the nearest possible class for elements. It works by finding the Euclidian distance between two points. One of these points is the source or the element that needs to be classified. The second point is selected from the given dataset and previous classifications. The algorithm calculates the Euclidian distance of the given data points to predict where an unknown point must lie [22].

3. MACHINE LEARNINGIN FOG COMPUTING

Fog is the essential player in the next generation of latency-free network. Introducing intelligence in the Fog environment can bring immense scalability to the network. Introducing intelligence in Fog implies integration of learning capabilities into the Fog network[27]. Context awareness, real-time response and resource optimization are just few of the benefits. Two broad categories of this intelligence could be device-driven approach and human-driven approach [28].
Device-driven Approach: This approach implies equipping devices with smarter and more robust devices. It could include computing, sensing, storage and communication. This approach is more dominant in Fog networks where sensors are varied and their data is diverse.
Human-driven Approach: This approach implies embedding human intelligence into machines, thereby improving performance. The key idea is to train the network through human behavior. Considerable works have been done in this regard. It involves scheduling the household appliances or enhancing the quality of life [28].

The input from the edge devices moves to the data centers for immediate storage and capturing. The core of data analysis relies heavily on Machine learning and its numerous algorithms. To embed thinking into a machine points to traditional and modern approaches of machine learning. These may include brute force and greedy K-nearest neighbor, neural networks, support vector machines or nature-inspired genetic algorithms. Moreover, advanced hardware such as GPU or embedded platforms can locally perform data-analytics. Cloud is then specialized for specific tasks such as data training for huge datasets[20].

Advantages:
A considerable point here to understand many potential benefits that come along this architecture. Efficient energy consumption is one such benefit. Fog devices are resource-constrained and applications such as image and voice processing are energy intensive. Catering the requirements of these devices over the edge of the network could mean task offloading, thereby conserving the precious bandwidth. Scheduling job sequence in this scenario is still a challenge.
Notable advantages are given below[28]:

3.1 Reduced Latency

Introducing Fog devices concludes that traffic on the network would be entertained locally. Some IoT devices such as real-time health monitoring system are time and resource intensive. If these Fog nodes are embedded with intelligent decision-making systems the capability of the system and performance could increase many folds[28].

3.2 Bandwidth Conservation

With localized preprocessing and storage of data near the edge of the network, huge volume of data can be addressed locally. This essentially frees up the constrained bandwidth since the amount of user data has increased exponentially over the years. Smart machine learning algorithms can provide the much-needed solution[28].
Offloading the computational tasks adds another strength to the Network. It provides means for preserving the battery life in pervasive devices. Specific machine learning algorithm namely Neuro-fuzzy logic using Neural networks are highly applicable in this perspective[5].

3.3 Improved Network Availability

Whenever a device moves from one region to another in Fog or Cloud environment, it experiences disruption and connectivity glitch. Intelligent machine learning algorithms should enable self-organization and co-operation to ensure smooth and uninterrupted services.
Security and Privacy: Self-organizing and learning also provides robust protection against malicious attacks. It is because network is able to identify and analyze data locally before propagating it to the network.

3.4 Optimization

Enabling intelligence in nodes can be beneficial in reducing the design complexity of a given scenario. Optimal Resource utilization prevents conflicts. It provides an acceptable tradeoff between accuracy and time complexity [12],[28].

4. APPLICATIONS OF MACHINE LEARNING ALGORITHMS IN FOG COMPUTING

Few major areas where highest impact is expected are as follows:

4.1 Health

The impact of machine learning and fog computing is huge in the health sector. Various applications are being proposed in this regard. One such application is IoTPulse. This project predicts the probability of a person being a drug addict. It utilizes machine-learning algorithms deployed in Fog Networks [22].
Another study revolves around the detection of a cardiac arrest using the wearable sensors[4]. The introduction of an efficient machine learning algorithm namely ROC analysis has helped in identifying most significant clinical parameters to get cardiac arrest [4].

4.2 Smart Grid

An application of Fog computing and machine learning lies in smart grids and energy management. A grid is an arrangement of various networks ranging from micro grids to international levels. Its applications include power generation and distribution systems, control systems for balancing and energy production. Due to highly dynamic and volatile nature of this application, an efficient fault tolerant mechanism is also needed. The inclusion of self-learning in this arrangement would enhance the error-detection and correction capabilities of the system[2].

4.3 Security

The Cloud services come with the issues of latency and connectivity. Moreover, due to performance requirements, the real-time implementation of effective security policy is difficult to achieve. In this perspective, the Fog devices can play a major role in offloading the traffic, conserving bandwidth and enforcing security. The Fog architecture, boasting horizontal connectivity has dual effects: it reduces storage space and latency issues. Fog nodes are efficient at handling widespread and massive information. The detection of attacks can be further controlled using the Machine learning algorithms. Moreover, the highly parallel nature of the Fog eases the mathematical modelling and calculations[3],[13].

4.4 Data Aggregation

 The synthesis, analysis and study of data aggregation produced an array of innovative benefits. But the issues related to data piracy and privacy associated with sensitive,

unhandled user data is a key concern[6].Intrusion Detection systems using Fog nodes are widely under development and research. Intrusion detection systems form the keystone of security systems. The resource constrained features of the Fog require specialized machine learning algorithms[7].

4.5 CPU Job Scheduling

Efficient scheduling of the jobs that need to be catered by the Fog environment is a research question. An experiment conducted showed that machine learning algorithms enhance CPU scheduling problem. This is an ideal requirement in the Fog networks. It enhances the operational capacity of the network and upgrades its performance. A significant algorithm here is Bee swarm algorithm that captures the essence of Bees life. This particular algorithm finds the optimal tradeoff between CPU execution time and allocated memory. An enhanced version of this algorithm can be utilized in network optimization [5].

4.6 Smart Cities

Smart city is also termed as "sustainable city" or self-dependent city. It implies a real-world scenario of enabling a city to be smart enough to identify its own needs. Geo-distribution is a major characteristic for the big data generated by smart cities. This huge data can only be handled using a massive collection of nodes such as in Fog. The data analysis for such varied and diverse data can be achieved affectively by using an effective machine learning algorithm[18].

4.7 Connected Vehicles

Fog computing along with machine learning is also incorporated in the smart vehicles. Efficient and effective machine learning algorithms to minimize handover optimization is possible [17]. Vehicular Fog simulator and Recurrent neural network can be used to minimize the latency and handover delay [17].

4.8 Network Connectivity and Detection

The next generation of Internet revolves around highest internet connectivity and broadband i.e. 5G. These networks are extremely dynamic and require great bandwidth. Due to the heterogeneous nature of these platforms, many challenges for network planning, deployment and operations are expected. Concurrently, the huge volumes of data being generated are a big question for scalability. Such challenges require dynamic resource provisioning and intelligent approaches. Fog offers great exploration of these resources. The dynamic and complex nature of future networks call for an intelligent and flexible structure[26]. This structure is clearly embedded in the Fog. Numerous machine-learning algorithms implemented here is include linear programming (LP), ant colony optimization (ACO) and genetic algorithm (GA) [12]. Natural language processing techniques are being used for enhancing the efficiency of wireless network

operations[27].The tables 5-11 below signify the application of each machine learning algorithm.

Table 5: Random Forest[2]

Domain	Functional Category	Task
Smart Grid	Forecasting/Prediction/Regression	demand side load forecasting/Price forecasting
	Anomaly/Fault Detection	Power record faults
Transport	Pattern Recognition/Clustering	Behavior/Event Recognition
Manufacturing	Anomaly/Fault Detection	Tooling wear/ Errors detection

.

Table 6:ANN[2]

Smart Grid	Forecasting/Prediction/Regression	Electrical Power prediction, load forecasting
Transport	Pattern Recognition/Clustering	Behavior/Event Recognition
	Forecasting/Prediction/Regression	traffic flow features
		road□side CO and NO2 concentrations estimation
		travel time prediction
	Classification	obstacle detection and recognition
		Image Processing
Manufacturing	Forecasting/Prediction/Regression/ optimization	Energy Consumption/ Process parameters optimization

Table 7:SVM[2]

Smart Grid	Forecasting/Prediction/Regression	Price Prediction
		Electrical Power prediction,
	Anomaly/Fault Detection	Non□Technical Loss detection
		Blackout Warning
		Power Line Attacks
Transport	Classification	Unintentional vehicle lane departure prediction
		Obstacles classification

	Pattern Recognition/ Clustering	Behavior/Event Recognition
	Anomaly/Fault Detection	Mechanism Failure
	Forecasting/Prediction/ Regression	Travel time prediction
Manufacturing	Forecasting/Prediction/ Regression	Machine Maintenance
	Design / Configuration	Feature Design; Production Processing
	Anomaly/Fault Detection	Quality Control
Smart Home	Pattern Recognition/ Clustering	Activity recognition

Table 8: Bayesian Network[2]

Transport	Classification	Event and behavior detect
Smart Grid	Anomaly/Fault Detection	Non□technical losses and fault detection
Manufacturing	Anomaly/Fault Detection	Fault detection in the production line
	Forecasting/Prediction/Re gression	Tool wear prediction/Energy consumption prediction

Table 9: Evolutionary Computing[2]

Smart Grid	Optimization/ Forecasting/Prediction	Short Term load forecasting

Table 10: Q□ learning□based algorithm[2]

Smart Grid	Optimizatio n	Aided Optimal Customer Decisions for an Interactive Smart Grid
Transport	Optimizatio n	the road latent cost

Table11: KNN[2]

Smart Grid	Optimization/ Forecasting/Prediction	Short Term load forecasting

5. FUTURE RESEARCH

Machine learning algorithms play a major role in enhancing the performance and efficiency of the system. The paper discussed various applications of Fog computing using machine learning. The various algorithms for machine learning were also discussed. Enabling intelligence in fog computing using machine learning offers huge ground for research and implementation[28] .

An important research question is to address device heterogeneity.

An example of computational offloading utilizing machine learning further covers research gap [29]. Considering the heterogenous nature of multiple devices, the choice of optimal machine learning algorithm to facilitate mobility and computational offloading in a specific application presents potential research area[29].

Considering the heterogenous nature of multiple devices, the choice of optimal machine learning algorithm to facilitate mobility and computational offloading in a specific application presents potential research area [29].

Similar research gap exits in fog resource allocation in blockchain networks [30]. An optimal machine learning algorithm that maximizes resource allocation and the arrival pattern of requests is an unanswered question [30].

Another unexplored paradigm is the handling of vast amount of data in a Wi-Fi environment both in delayed and non-delayed transmissions [31]. The optimization of Genetic algorithm to best suit varied architecture for differentiated data is still to be answered [31].

The debate of optimal resource allocation and management for energy conservation is still open [32]. The quest for ideal design that promises best resource allocation for energy conservation in contrary to existing designs is still to conclude[32].

Security and access control to meet different stakeholders is a research area in itself. The overwhelming amount of data and its variety, its storage and access methods and the choice of optimal machine learning algorithm to ensure optimal performance are still unanswered [33].The fine grained control provided by the Fog network and best-fit machine learning algorithm are major design concerns [33].

Another research area is the load balancing in a dynamic Fog environment [34]. The challenge of load balancing utilizing Q-learning with greedy algorithm can be enhanced further [34]. The optimization techniques to achieve lower processing delay and less machine dependency can be studied further[34].

The Fog to Cloud communication is another research gap. The design techniques to minimize latency and speedup the Fog-Cloud communication are being researched [35].This is the major architectural design for Internet of Things [35].The ideal machine learning algorithm for data caching at fog nodes would reduce network latency. Machine learning techniques to predict future needs are majorly unanswered[35].

As previously stated, one way to view Fog Networks is to envision it as a Cyber-physical system. There are numerous applications of this visualization [36].They include smart buildings, power grids and connected vehicles. The design concerns for individual realization are largely developed through machine learning models. Yet the optimal choice of a specific machine learning algorithm based on a particular scenario is still unanswered.The design aspects such as energy efficiency, computational offloading and reduced latency are increasingly important research areas [36].

Scalability and sustainability in a Fog architecture is also one among many other design concerns. Due to horizontal, wide-spread,layered architecture of Fog nodes, network authentication and resource provisioning are an open challenge [37].The common model that caters device heterogeneity is lacking.The maximum number of devices that can be connected to a Fog Nodedepends on aspecific scenario.The

Quality of Service (QoS) can be seriously compromised if the number of connected devices exceed the threshold.Moreover, Fog nodes are acting as local data processing units.Hence the need for change the functionality as per the scenario needs to be embedded in the Fog nodes.Thus,Fog nodes virtualization is another open research question[37].

Task scheduling in a CPS is also an important research direction [38] .Considering Fog as an expected framework for CPS, task scheduling to fulfil QoS is an unanswered question. An implementation of task scheduling utilizing moth's behavior offers a promising solution[38]. The future direction points to integration of task-scheduling with load balancing[38].

The optimal design for load balancing and task scheduling is greatly applicable in Vehicle Adhoc Networks (VANETs). The load balancing approach focuses on finding the optimal path from the source to the destination, better network partitions and higher transmission rates. Intelligent decisions can be made for task scheduling if machine learning algorithms are properly utilized. In such a scenario, the consumer satisfaction and data privacy take the central stage. Efficient buffer design, scalability and security are sizable research gaps[39].

6.CONCLUSION

The applications of machine learning algorithms in Fog computing are numerous. there exist major applications where this fusion can be game changing. This design paradigm can be effective in designing the optimal solution for Internet of Things. Deploying traditional machine learning techniques to make fog nodes intelligent can greatly enhance the IoT networks [40].Optimal resource provisioning, security, task scheduling and many other important design parameters can be addressed by integrating machine learning with Fog devices [40].Minimal computational latency, real-time analysis and addressing device heterogeneity in Fog Networks can be addressed using machine learning algorithms[41].

REFERENCES

1. U. Y. Khan and T. R. Soomro, "**Envisioning Internetnet of Things using Fog Computing**," *International Journal of Advanced Computer Science and Applications,* vol. 9, no. 1, pp. 441-448, 2018.

2. X. Fei, N. Shah, N. Verba, K.-M. Chao, V. Sanchez-Anguix, J. Lewandowski, A. James and Z. Usman, "**CPS data streams analytics based on machine learning for Cloud and Fog Computing: A survey**," *Future Generation Computer Systems,* vol. Volume 90, pp. 435-450, 2019.

3. U. Y. K. a. T. R. Soomro, "**Applications of IoT: Mobile Edge Computing Perspectives,**" in *12th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)*, Karachi,

Pakistan, 2018 .

4. P. . M. Kumar and U. D. Gandhi, "**A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases**," *Computers & Electrical Engineering,* vol. Volume 65, pp. Pages 222-235, 2018.

5. F. Haouari, . R. Faraj and J. M. AlJa'am, "**Fog Computing Potentials, Applications, and Challenges**," in *International Conference on Computer and Applications (ICCA)*, Beirut, 2018.

6. D. Kimovski, H. Ijaz, N. Saurabh and R. Prodan, "**Adaptive Nature-Inspired Fog Architecture**," in *IEEE 2nd International Conference on Fog and Edge Computing (ICFEC)*, Washington, DC, 2018.

7. X. An, X. Zhou, X. Lü, F. Lin and L. Yang, "**Sample Selected Extreme Learning Machine Based Intrusion Detection in Fog Computing and MEC**," *Wireless Communications and Mobile Computing,* vol. vol. 2018, pp. 1-10, 2018.

8. A. A. Alli and M. . M. Alam, "**SecOFF-FCIoT: Machine learning based secure offloading in Fog-Cloud of things for smart city applications,**" *Internet of Things,* vol. Volume 7, September 2019 2019.

9. M. Lavassani, S. Forsström, U. Jennehag and T. Zhang, "**Combining Fog Computing with Sensor Mote Machine Learning for Industrial IoT**," *Sensors,* Vols. vol. 18, no. 5, no. vol. 18, no. 5, p. 1532, May 2018.

10. M. Yang, T. Zhu, . B. Liu, . Y. Xiang and W. Zhou, "**Machine Learning Differential Privacy With Multifunctional Aggregation in a Fog Computing Architecture**," *IEEE Access,* vol. Volume: 6, pp. 17119 - 17129, 20 March 2018.

11. D. Wu, S. Liu, L. Zhang, J. Terpenny, R. X. Gao, T. Kurfess and J. A. Guzzo, "**A fog-computing based framework for process monitoring and prognsis in cyber-manufacturing**," *Journal of Manufacturing Systems,* Vols. Volume 43,, no. Part 1,, pp. Pages 25-34,, 2017.

12. Z. Tang, . X. Zhou, F. Zhang, W. Jia and W. Zhao, "**Migration Modeling and Learning Algorithms for Containers in Fog Computing**," *IEEE Transactions on Services Computing,* Vols. vol. 12, no. 05, no. vol. 12, no. 05, pp. pp. 712-725, 2019.

13. M. Moh and . R. Raju, "**Machine Learning Techniques for Security of Internet of Things (IoT) and Fog Computing Systems**," in *2018 International Conference on High Performance Computing & Simulation (HPCS)*, Orleans, 2018.

14. I. Alrashdi, A. Alqazzaz, E. Aloufi, R. Alharthi, M. Zohdy and H. Ming, "**AD-IoT: Anomaly Detection of IoT Cyberattacks in Smart City Using Machine Learning**," in *IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, NV, USA,, 2019.

15. Z. Shen, . K. Yokota, J. Jin, . A. Tagami and T. Higashino,

"**In-network Self-Learning Algorithms for BEMS Through a Collaborative Fog Platform**," in *IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA),*, Krakow, 2018.

16. . D. Rahbari and . M. Nickray, "**Task offloading in mobile fog computing by classification and regression tree.**," *Peer-to-Peer Networking and Applications,* vol. 13, p. 104–122, 2019.

17. S. Memon and M. Maheswaran, "**Using machine learning for handover optimization in vehicular fog computing,**" in *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, Limassol Cyprus, April 2019.

18. B. Tang , Z. Chen, G. Hefferman, P. Shuyi , T. Wei, H. He and Q. Yang, "**Incorporating Intelligence in Fog Computing for Big Data Analysis in Smart Cities**," *IEEE Transactions on Industrial Informatics,* Vols. vol. 13, no. 5, no. vol. 13, no. 5, pp. pp. 2140-2150, Oct. 2017.

19. Y. Tu, Y. Ruan, S. Wagle, C. G. Brinton and C. Joe-Wong, "**Network-Aware Optimization of Distributed Learning for Fog Computing**," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*, Toronto, ON, Canada, 2020.

20. K. H. Abdulkareem, . M. . A. Mohammed, S. Gunasekaran, A. Mutlag, . A. M. Salama and N. Salih Ali, ""**A Review of Fog Computing and Machine Learning: Concepts, Applications, Challenges, and Open Issues,**," *IEEE Access,* vol. vol. 7, no. vol. 7, pp. pp. 153123-153140, 2019.

21. J. C. Guevara, R. . d. S. Torres and N. L. d. Fonseca, "**On the classification of fog computing applications: A machine learning perspective**," *Journal of Network and Computer Applications,* vol. Volume 159, no. Volume 159, pp. 1-10, 2020.

22. A. Dhillon, A. Singh, H. Vohra, C. Ellis, B. Varghese and S. Gil, "**IoTPulse: machine learning-based enterprise health information system to predict alcohol addiction in Punjab (India) using IoT and fog computing**," *Enterprise Information Systems,* pp. 1-34, 21 Sep 2020.

23. E. Balevi, and . R. Gitlin, "**Unsupervised machine learning in 5G networks for low latency communications,**" in *2017 IEEE 36th International Performance Computing and Communications Conference (IPCCC)*, University of South Florida, 2017.

24. A. MelloukImag, S. Batim and S. Zeadally, "**Fog computing job scheduling optimization based on bees swarm,**" *Enterprise Information Systems,* Vols. Volume 12, 2018, no. Volume 12, 2018, 10 Apr 2017.

25. A. Abeshu and N. Chilamkurti, "**Deep Learning: The Frontier for Distributed Attack Detection in Fog-to-Things Computing**," *IEEE Communications Magazine,* Vols. vol. 56, no. 2, no. vol. 56, no. 2, pp. pp. 169-175, Feb. 2018.

26. A. Al-Ghraibah, M. Algharibeh, W. AlMohtasib and M. Altayeb, "**Classification of Long Bone X-ray Images using New features and Support Vector Machine,**"

Amani Al-Ghraibah et al., International Journal of Advanced Trends in Computer Science and Engineering, vol. 10, no. 3, pp. 1494-1500, 12 6 2021.

27. T. . E. Bogale, X. Wang and L. B. Le, "**Machine Intelligence Techniques for Next-Generation Context-Aware Wireless Networks,**" *ITU Special Issue N.1 The impact of Artificial Intelligence (AI) on communication networks and services,* 12 Jan 2018.

28. Q. . D. La, M. V. Ngo, T. Q. Dinh, T. Q.S. Quek and H. Shin, "**Enabling intelligence in fog computing to achieve energy and latency reduction,**" *Digital Communications and Networks,* Vols. Volume 5, Issue 1, no. Volume 5, Issue 1, pp. 3-9, 2019.

29. S. Zhou, . S. Y, Z. Jiang and Z. Niu, ""**Exploiting Moving Intelligence: Delay-Optimized Computation Offloading in Vehicular Fog Networks,**" *IEEE Communications Magazine,* Vols. vol. 57, no. 5, no. vol. 57, no. 5, pp. pp. 49-55, May 2019.

30. N. C. Luong, Y. Jiao, P. Wang, D. Niyato, D. I. Kim and Z. Han, "**A Machine-Learning-Based Auction for Resource Trading in Fog Computing,**" *IEEE Communications Magazine,* Vols. vol. 58, no. 3, no. vol. 58, no. 3, pp. pp. 82-88, March 2020.

31. George-Eduard Zaharia, T.-A.-I. Şoşea, R.-I. Ciobanu and C. Dobre, "**Machine learning-Based traffic offloading in fog networks**," *Simulation Modelling Practice and Theory,,* vol. Volume 101, 2020,.

32. W. Shafik, , . s. M. Matinkhah and S. Mamman, "**Network Resource Management Drives Machine Learning: A Survey and Future Research Direction,**" *Journal of Communications Technology, Electronics and Computer Science,* vol. Issue 30, no. Issue 30, 2020.

33 A. S. M. Kayes, R. Kalaria, I. H. S. Sarker, I. H. Sarker, M. S. Islam and P. A. Watters, "**A Survey of Context-Aware Access Control Mechanisms for Cloud and Fog Networks: Taxonomy and Open Research Issues**," *Sensors,* vol. vol. 20, no. 9, p. 2464, Apr. 2020.

34. J. Baek, G. Kaddoum, S. Garg, K. Kaur and V. Gravel, "**Managing Fog Networks using Reinforcement Learning Based Load Balancing Algorithm**," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, Marrakesh, Morocco, 2019.

35. RIYA, . G. NITIN and Q. XIN, **Data Caching at Fog Nodes Under IoTNetworks: Review of Machine Learning,** techrxiv.org, 2020.

36. J. Moura and D. Hutchison, "**Fog Computing Systems: State of the Art, Research Issues and Future Trends**," *Journal of Network and Computer Applications,* vol. Volume 169, no. Volume 169, 2020.

37. R. Z. Naeem, S. Bashir, M. . F. Amjad, H. Abbas and H. Afzal , "**Fog computing in internet of things:Practical applications and future directions**," *Peer-to-Peer Networking and Applications,* vol. 12, no. 12, p. 1236–1262, 13 September 2019.

38. M. Ghobaei-Arani, A. Souri, F. Safara and N. Monire , "**An efficient task scheduling approach using**

moth-flame optimization algorithm for cyber-physical system applications in fog computing," *Trans Emerging Tel Tech.,* pp. 1-14, 4 September 2019 4 September 2019.

39. H. A. Khattak, S. U. Islam, I. U. Din and M. Guizani, "**Integrating Fog Computing with VANETs: A Consumer Perspective**," *IEEE Communications Standards Magazine,* vol. vol. 3, no. 1, pp. pp. 19-25, March 2019.

40. F. Hussain, . R. Hussain, S. . A. Hassan and . E. Hossain, "**Machine Learning in IoT Security: Current Solutions and Future Challenges**," *IEEE Communications Surveys & Tutorials,* vol. vol. 22, no. no. 3, pp. 1686-1721, thirdquarter 2020.

41. D. V and L. S. R, "**Intelligent Deep Reinforcement Learning Based Resource Allocation in Fog Network**," in *26th International Conference on High Performance Computing, Data and Analytics Workshop (HiPCW),*, Hyderabad, India, 2019.